

MASTER'S THESIS

Intuïtiviteit in Enterprise Architectuur modelleertalen: een vergelijkende empirische evaluatie

Meulema, G.B. (Gudrun Bliede)

Award date:
2021

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy

If you believe that this document breaches copyright please contact us at:

pure-support@ou.nl

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 06. May. 2023

Open Universiteit
www.ou.nl



Intuïtiviteit in Enterprise Architectuur modelleertalen: een vergelijkende empirische evaluatie

Intuitivity in Enterprise Architecture modelling languages: a comparative empirical evaluation

Opleiding:	Open Universiteit, faculteit Management, Science & Technology Masteropleiding Business Process Management & IT
Programme:	Open University of the Netherlands, faculty of Science Master of Business Process Management and IT
Cursus:	IM0602 Voorbereiden Afstuderen BPMIT IM9806 Afstudeeropdracht Business Process Management and IT
Student:	Gudrun Bliede Meulema
Identiteitsnummer:	
Datum:	24 januari 2020
Afstudeerbegeleider	Ben Roelens
Meelezer	
Derde beoordelaar	
Versie nummer:	V01
Status:	Definitief

Abstract

Om als organisatie succesvol te zijn, is het noodzakelijk dat het geheel van processen, procedures en protocollen wordt begrepen door alle lagen en over alle disciplines. De rol van enterprise architectuur is hierin groot. Hieruit wordt de strategie vertaald naar een organisatiemodel. Om dit op alle niveaus te visualiseren, wordt gebruik gemaakt van modelleertalen, waarbij makkelijk kunnen begrijpen, intuïtiviteit, een must is. Een van de talen die hiervoor is ontwikkeld, is de proces-goal alignment techniek (PGA)

Dit onderzoek heeft als doel een methodiek te vinden om de intuïtiviteit van de oorspronkelijke en verbeterde versie van de PGA-techniek vergelijkend te evalueren en zo te toetsen of de verbetering in intuïtiviteit van de PGA techniek zich ook daadwerkelijk heeft voorgedaan.

Hiertoe is een experiment opgezet met een gemengd ontwerp. Respondenten kregen naast demografische vragen een viertal experimentele taken voorgelegd.

Om het begrip “intuïtiviteit” meetbaar te maken en zodoende vergelijkend te kunnen is gebruik gemaakt van de begrippen “interpretatieve effectiviteit” en “interpretatieve efficiëntie”, beiden als objectieve meeteenheden voor de eerste drie taken. De “ervaren intuïtiviteit” is als subjectieve meeteenheid aangehouden in de vierde en laatste taak.

Een belangrijke factor binnen de opzet van dit onderzoek is de wereldwijde pandemie gebleken. Hierdoor is noodgedwongen een online onderzoek uitgevoerd. Dit heeft tot gevolg gehad dat de omstandigheden waaronder de respondenten hebben deelgenomen in potentie beïnvloed konden worden door externe factoren.

Een tweede beperking is de wijze van persoonlijke reflectie gebleken. Door de keuze uit twee elementen voor te leggen, waarbij de motivatie voor de keuze facultatief was, is keuze voor “de minst slechte” niet uit te sluiten.

De resultaten geven aan dat niet alle verschillen als significant kunnen worden aangemerkt. Daarnaast bleken verschillen wisselend van aard. Hiermee kan niet onomstotelijk worden vastgesteld dat de verbeterde PGA-notatie beter scoort dan de oorspronkelijke PGA-notatie.

De twee belangrijkste verklaringen hiervoor kunnen zijn dat de verschillen tussen de oorspronkelijke notatie en de verbeterde notatie eenvoudigweg te klein zijn. Een tweede mogelijke verklaring is dat wellicht de gebruikte methode niet of niet volledig geschikt is om de intuïtiviteit van de notatie te bepalen. Vervolgonderzoek zal zich dan ook moeten richten op het toetsen van de methode en het vergroten van de verschillen tussen twee (varianten van) modelleertalen.

Sleutelbegrippen

PGA, Intuïtiviteit, Enterprise Architectuur, DSML, interpretatief

Samenvatting

In dit onderzoek is gepoogd een antwoord te krijgen op de vraag hoe intuïtief een Enterprise architectuurmodelleertaal is en deze te evalueren aan de hand van de Proces-Goal- Alignment techniek (PGA) in haar oorspronkelijk versie (PGA-1) en de verbeterde versie (PGA-2).

Uit literatuuronderzoek is gebleken dat specifiek voor dit onderzoek intuïtief betekent: zonder voorkennis kunnen begrijpen en toepassen van de bij de taal horende notaties.

Om een subjectief begrip als intuïtiviteit objectief meetbaar te maken is, gebruikmakend van eerder onderzoek gericht op het evalueren van modelleertalen op intuïtiviteit, gekozen voor toepassing van de begrippen “*interpretatieve effectiviteit*” en “*Interpretatieve efficiëntie*”, eenheden die de juiste antwoorden op associatieve -, begrips- en herinneringstaken alsmede de geaggregeerde antwoorden over de te toetsen elementen respectievelijk de daarbij behorende besteedde tijd per antwoord meten.

Daarnaast is respondenten gevraagd hun voorkeur aan te geven voor een van beide varianten en hierbij hun keuze te motiveren, die de “ervaren intuïtiviteit” weergeeft. Hierbij is weliswaar sprake van een subjectieve meeteenheid, echter uit onderzoek is gebleken dat de persoonlijke reflectie binnen evaluatiemethoden van modelleertalen onmisbaar is.

Noodgedwongen door de uitbraak van COVID-19 is gekozen voor een volledig onlineonderzoek. Dit maakt met name de fysieke omgeving waarin de respondent deelneemt aan het onderzoek niet beïnvloedbaar, terwijl deze fysieke omgeving omgekeerd wel invloed kan hebben op de onderzoeksresultaten d.m.v. afleiding voor de respondent.

Uit de resultaten is gebleken dat, hoewel er verschillen tussen beide varianten van de PGA-notaties konden worden geconstateerd, deze verschillen statistisch overwegend niet significant bleken. Dit betekent dat toeval in de resultaten niet kon worden uitgesloten. Desalniettemin bleek dat de scores in uitvoerende taken voornamelijk wezen op een betere score voor de oorspronkelijke versie van de taal, terwijl uit de reflectie door de respondenten duidelijk een voorkeur voor de verbeterde variant van de PGA-taal bleek.

Gezien het feit dat de resultaten overwegend niet significant zijn en daardoor toeval niet uitgesloten kan worden, kunnen de gestelde hypothesen niet onomstotelijk bevestigd of ontkracht worden. Evenwel biedt dit onderzoek wel mogelijkheden voor vervolgonderzoek.

Een mogelijke verklaring voor de resultaten van het onderzoek is dat de geconstateerde verschillen tussen de verschillende varianten te klein zijn om significant te zijn. Daarnaast zou het kunnen zijn dat de methode (deels) minder geschikt is om intuïtiviteit van een notatie te bepalen.

Hoewel op basis van dit onderzoek de tijd nog niet is aangebroken voor introductie van de PGA-techniek in de praktijk, is het wel zinvol om de methode in de praktijk te toetsen door enerzijds het aanbrengen van grotere verschillen tussen twee (varianten van) talen en anderzijds aan onderzoek uit andere wetenschapsgebieden zoals psychologie, hersenonderzoek in het algemeen en onderzoek naar dementie in het bijzonder. Daarnaast kunnen ook vanuit taalwetenschappen nieuwe inzichten worden gegenereerd.

Daarnaast heeft de keuze voor de wijze van reflecteren op de taal beperkingen met zich meegebracht. Omdat de keuze gemaakt moest worden uit twee symbolen kan niet worden uitgesloten dat respondenten hebben gekozen voor de “minst slechte keuze”. Ook heeft het vrijwillige karakter van de motivatie voor de keuze de persoonlijke reflectie meer het karakter gekregen van een gesloten vraag dan een open vraag. In vervolgonderzoek is het dan ook raadzaam om binnen de persoonlijke reflectie ofwel een verplichting tot beantwoording ligt, ofwel de persoonlijke reflectie door middel van interviews wordt uitgevoerd.

In de laatste plaats is dit onderzoek een van de eerste onderzoeken dat expliciet de intuïtiviteit van een modelleertaal vergelijkend evalueert onder professionals. In een dergelijk vroeg stadium is het altijd aan te bevelen om de methode zelf te toetsen.

Summary

This research has attempted to answer the question how intuitive an Enterprise architecture modeling language is and to evaluate it by means of the Process-Goal-Alignment technique (PGA) in its original version (PGA-1) and the improved version (PGA-2).

Literature research has shown that specifically for this research intuitive means: being able to understand and apply the notations associated with the language without prior knowledge.

In order to make a subjective concept such as intuitivity objectively measurable, using previous research aimed at evaluating modeling languages for intuitiveness, the choice was made to apply the concepts “interpretive effectiveness” and “interpretative efficiency”, units that measure the correct answers to associative-, comprehension- and recall tasks as well as the aggregated answers on the elements to be tested and the time spent per answer.

In addition, respondents were asked to indicate their preference for one of the two variants and to motivate their choice, which reflects the “perceived intuitiveness”. Although this is a subjective measuring unit, research has shown that personal reflection within evaluation methods of modeling languages is indispensable.

Out of necessity, due to the outbreak of COVID-19, a full online research was chosen. The particular problem is that the physical environment in which the respondent participates cannot be influenced by the researcher, while this physical environment, conversely, can influence the research results by means of distraction for the respondent.

The results showed that, although differences could be observed between the two variants of the PGA notations, these differences were statistically largely insignificant. This means that coincidence could not be excluded from the results. Nevertheless, it turned out that the scores in executed tasks mainly indicated a better score for the original version of the language, while the reflection by the respondents clearly showed a preference for the improved variant of the PGA language.

Given that the results are largely insignificant and therefore coincidence cannot be ruled out, the hypotheses made cannot be irrefutably confirmed or invalidated. However, this research does offer opportunities for further research.

A possible explanation for the results of the study is that the differences observed between the different variants are too small to be significant. In addition, it could be that the method is (partly) less suitable for determining the intuitiveness of a notation.

Although, based on the results of this research, the time has not yet come for the introduction of the PGA technique in practice, it is useful to test the method in practice by on the one hand making larger differences between two (variants of) languages and on the other hand to research from other scientific fields such as psychology, brain research in general and research into dementia in particular. In addition, new insights can also be generated from linguistics.

The choice of how to reflect on the language has brought its limitations. Because the choice had to be made between two symbols for the same element, it cannot be ruled out that respondents chose the “least bad choice”.

The voluntary nature of the motivation for the choice has also given the personal reflection more the character of a closed question than an open question. In follow-up research it is therefore

advisable to either have an obligation to fill in the motivation for the respondent's choice within the personal reflection, or to carry out the personal reflection by means of interviews.

Finally, this research is one of the first studies to explicitly evaluate the intuitiveness of a modeling language comparatively among professionals. At such an early stage it is always advisable to test the method used.

Inhoudsopgave

Abstract	iii
Sleutelbegrippen	iii
Samenvatting	iv
Summary	vi
Inhoudsopgave	viii
1. Introductie	1
1.1. Achtergrond	1
1.2. Gebiedsverkenning	2
1.3. Probleemstelling	3
1.4. Opdrachtformulering	3
1.5. Motivatie/ relevantie	4
1.6. Aanpak in hoofdlijnen	5
2. Theoretisch kader	5
2.1. Onderzoeksaanpak.....	5
2.2. Uitvoering.....	6
2.3. Resultaten en conclusies.....	8
2.3.1. Resultaten	8
2.3.2. Conclusie	14
2.4. Doel van het vervolgonderzoek	15
3. Methodologie.....	16
3.1. Conceptueel ontwerp: keuze van onderzoeksmethode(n)	16
3.2. Technisch ontwerp: uitwerking van de methode	16
3.2.1. Variabelen en maatstaven	16
3.2.2. Onderzoeksontwerp	17
3.2.3. Experimentele taken en wijze van toetsing	17
3.2.4. Hypothesen	21
3.2.5. Respondenten	21
3.3. Gegevensanalyse.....	22
3.4. Reflectie t.a.v. validiteit, betrouwbaarheid en ethische aspecten	22
4. Resultaten	25
4.1. Participatie	25

4.2.	H1: Interpretatieve effectiviteit van de PGA-2 notatie voor element X tijdens taak Y > Interpretatieve effectiviteit PGA-1 notatie voor element X tijdens taak Y.....	25
4.2.1.	Notatie-associatie	26
4.2.2.	<i>Comprehension</i> -taak	27
4.2.3.	Recall-taak.....	27
4.3.	H2: Interpretatieve efficiëntie PGA-2 notatie voor element X tijdens taak Y < Interpretatieve efficiëntie PGA-1 notatie voor element X tijdens taak Y	28
4.3.1.	Notatie-associatie	29
4.3.2.	<i>Comprehension</i> -taak	29
4.3.1.	<i>Recall</i> -taak.....	30
4.4.	H3: Ervaren intuïtiviteit PGA-1 < Ervaren intuïtiviteit PGA-2.....	30
4.5.	Geaggregeerde scores	31
5.	Discussie, conclusies en aanbevelingen.....	33
5.1.	Discussie – reflectie.....	33
5.1.1.	Validiteit.....	33
5.1.2.	Betrouwbaarheid	33
5.1.3.	Interpretatieve effectiviteit	34
5.1.4.	Interpretatieve efficiëntie.....	35
5.1.5.	Ervaren intuïtiviteit	35
5.2.	Conclusies	36
5.2.1.	Conclusies Hypothesen	36
5.3.	Aanbevelingen voor de praktijk.....	37
5.4.	Aanbevelingen voor verder onderzoek.....	38
	Referenties	39
	Bijlage 1: Notatie Associatie	40
	Bijlage 2: Casus voor Comprehension Questions	41
	Bijlage 3: Tabellen Recallquestions.....	43
	Bijlage 4: Statistische testen individuele dataset.....	44
	Bijlage 5: Statistische testen gecombineerde dataset	70
	Bijlage 6: Testen geaggregeerde scores	94

1. Introductie

1.1. Achtergrond

Enterprise architectuur (EA) neemt een belangrijke plaats in bij grotere organisaties. Toch blijkt er nog steeds niet altijd duidelijk te zijn wat EA nu eigenlijk inhoudt. Lankhorst et al (2009, p 3) beschrijven EA als :

“Een samenhangend geheel van principes, methoden en modellen die worden gebruikt in het ontwerp en de realisatie van organisatorische structuur, business processen, informatiesystemen en infrastructuur van een organisatie”

Uit de verschillende elementen in de definitie, wordt duidelijk dat EA zich begeeft op verschillende hiërarchische niveaus, te weten: strategisch, tactisch en operationeel. Deze hebben vooral betrekking op de business laag. Deze wordt ondersteund door de applicatie - en de daaronder liggende technologische laag. Deze lagen kunnen een op een worden vertaald naar de domeinen bedrijfsvoering (i.e. business laag), informatievoorziening (i.e. applicatie laag) en technologiedomein (i.e. technologische laag) en zijn onlosmakelijk met elkaar verbonden.

Eenzijds is er de strategie die moet worden vertaald naar een zowel tactisch als operationeel uitvoerbaar geheel van protocollen, richtlijnen en processen en anderzijds is daar de IT die dat uitvoerbare geheel optimaal zal moeten ondersteunen. In dit proces, waarin business en IT met elkaar in lijn moeten worden gebracht, is de rol van EA onmiskenbaar groot.

Om succesvol te kunnen zijn is het nodig dat het organisatieontwerp vanuit EA “landt” in alle lagen van de organisatie. Binnen het domein van EA wordt gebruik gemaakt van generieke modelleertalen, de *General Purpose Modelling languages* (GPML) en *Domain Specific Modelling languages* (DSML) om op verschillende niveaus het organisatieontwerp te visualiseren .

Een DSML wordt door Frank (2014, p. 320) omschreven als *“een modelleertaal die bedoeld is om gebruikt te worden in een specifiek domein”* en *“.. is gebaseerd op concepten die werden gereconstrueerd uit technische termen die werden gebruikt in het respectievelijke domein”*. Hierbij benoemt hij ook duidelijk de voordelen van DSML ten opzichte van GPML: de bevordering van productiviteit bij modelbouwers en de bevordering van systeemintegriteit en een beter medium voor communicatie met potentiële gebruikers, als gevolg van de duidelijke syntax en semantiek van de domein-specifieke termen.

Hoe intuïtiever de modelleertaal, hoe groter de kans dat het samenhangende geheel door alle niveaus heen begrepen wordt. Maar, wat verstaan we onder intuïtiviteit? In de volksmond wordt intuïtie beschreven als iets wat je direct weet, zonder erover na te (hoeven) denken of beredeneren. Jošt, Huber, Heričko, and Polančič (2016) definiëren in hun onderzoek naar de intuïtiviteit van procesmodellen intuïtiviteit als iets, een notatie, dat zonder voorkennis of training kan worden begrepen. Daarnaast blijkt uit het onderzoek van Roelens en Bork (2020) dat het begrip ook gaat om de taal te kunnen toepassen zonder voorkennis of opleiding.

1.2. Gebiedsverkenning

Onderzoek naar de afstemming van de organisatiestrategie met de business-architectuur heeft geleid tot de ontwikkeling de *Process – Goal Alignment* techniek, afgekort tot PGA (Roelens, Steenacker, & Poels, 2019). Deze DSML is erop gericht om verschillende hiërarchische niveaus binnen de EA met elkaar te integreren, waardoor er een holistisch beeld ontwikkeld kan worden dat rekening houdt met alle stakeholders.

Juist omdat het om verschillende hiërarchische niveaus gaat, is begripsvorming en gebruiksgemak essentieel. De mate van direct begrip van de symbolen van de PGA-taal en het gebruiksgemak, geeft de intuïtiviteit van de taal aan. Om de intuïtiviteit van de bestaande PGA notatie te evalueren is een onderzoek uitgevoerd volgens een door Bork, Schruffer, and Karagiannis (2019) specifiek ontwikkelde methode om intuïtiviteit van modelleertaal notaties te evalueren. Deze door hen geïntroduceerde methode is het vertrekpunt voor de methode in dit onderzoek. Na analyse van de onderzoeksresultaten werden voorstellen geformuleerd om de notatie te verbeteren (Roelens & Bork, 2020). Deze techniek bestaat uit de volgende fasen:

- *Fase 1 – term association*: elke deelnemer krijgt een lijst met termen, afkomstig uit de te evalueren modelleertaal. Vervolgens wordt elke individuele deelnemer gevraagd een grafische weergave te tekenen per term, die zij als meest intuïtieve weergave van de betreffende term beschouwen.

De analyse houdt de deconstructie van de grafische weergaven in op verschillende punten, zoals vorm, symbolen en kleur. Door de meest gebruikte visuele elementen te vergelijken met de huidige notatie kan inzicht worden verworven in mogelijke verbeteringen (van de intuïtiviteit) van de notatie.

- *Fase 2 – notation association*: elke deelnemer krijgt de huidige notaties binnen de PGA-taal gepresenteerd, met het verzoek hierbij tot drie associaties op te schrijven. Deze notaties verschillen van de termen die gebruikt zijn in de eerste fase van het experiment, om te voorkomen dat de resultaten van deze fase in het experiment worden beïnvloed.

De analyse houdt in dat voor de gegeven antwoorden moet worden berekend wat het percentage respondenten is dat een juiste associatie heeft gegeven. Dit geeft een goed beeld van de mate van intuïtiviteit van het gebruikte symbool.

- *Fase 3 – case study*: Er wordt onderzocht of de deelnemers in staat zijn om de PGA-concepten intuïtief te combineren tot een model. De casus moet hierbij zo kort en compact mogelijk zijn, waarbij de kwaliteit van het resultaat overeind blijft. Hiertoe zijn twee mogelijkheden:
 - De deelnemer krijgt een casus opgegeven (schriftelijk) en de beschikking over een modelleertool om zo de modelleertaak te kunnen vervullen.
 - Zijn er veel deelnemers of is er weinig tijd beschikbaar, dan kan gekozen worden voor een voorbeeldmodel met begripsvragen.

Bij de eerste variant richt de analyse zich op de semantische en syntactische juistheid van het uitgewerkte model. In variant twee richt de analyse zich op de (juistheid) van de gegeven antwoorden.

- Analyse & verbeterde notatie: Hierop volgt een afrondende fase waarin de deelnemers een feedbackformulier krijgen, waarop zij positieve en negatieve feedback kunnen geven en verbeter suggesties kunnen doen.

1.3. Probleemstelling

De PGA-modelleertaal heeft een participatief ontwerp. Dat maakt dat het ontwerp van de oorspronkelijke notatie afhankelijkheid is van menselijke creativiteit.

De afhankelijkheid van menselijke creativiteit geldt ook voor de uitvoering én evaluatie van de experimentele taken. Vanuit het eerder uitgevoerde experimentele onderzoek naar de intuïtiviteit van deze conceptuele modelleertaal (Roelens & Bork, 2020) is een verbeterde versie van de PGA-taal voortgekomen.

De intuïtiviteit van de PGA-taal die hieruit is voortgekomen kan niet worden aangetoond en verdient nader onderzoek. Door de initiële versie van de PGA-modelleertaal (hierna te noemen PGA-C0.1) te vergelijken met de voorgestelde verbeterde versie van de modelleertaal (PGA-C0.2), kan de mate van verbetering qua intuïtiviteit empirisch worden onderzocht.

De overkoepelende onderzoeksvraag die gesteld wordt is:

Welke methodiek kan worden ingezet om de oorspronkelijke- en verbeterde PGA-notatie vergelijkend te evalueren?

1.4. Opdrachtformulering

Uitgaande van de overkoepelende vraag zijn er sub-vragen die beantwoord moeten worden. Hiervoor zullen we meer inzicht nodig hebben in bestaande onderzoeksmethoden. Deze bestaande onderzoeksmethoden zijn immers beproefd en gevalideerd en bieden daarmee een wetenschappelijk onderbouwd raamwerk om het onderzoek waarop deze paper betrekking heeft vorm te geven. Daarnaast kan, door dit onderzoek te bouwen op eerder uitgevoerd onderzoek, kennis worden toegevoegd aan de bestaande wetenschappelijke inzichten over onderzoeksmethodieken die zijn gericht op het vergelijkend evalueren van modelleertalen.

Het onderzoek gaat, zoals gezegd, om het vergelijkend evalueren van een oorspronkelijke en verbeterde notatie van een modelleertaal. De vraag die dan ook beantwoord moet worden is de volgende:

1. Welke relevante methoden zijn er beschreven om modelleertalen te vergelijken en te evalueren?

Antwoord op deze vraag zal worden gevonden middels literatuuronderzoek.

Doel van het vergelijken van de oorspronkelijke notatie met de verbeterde notatie is enerzijds om te vast te kunnen stellen of de verbetering zich ook daadwerkelijk heeft voorgedaan en anderzijds om de mate van intuïtiviteit vast te kunnen stellen. Om de significantie van de verbetering objectief vast

te kunnen stellen, zal een subjectief begrip als intuïtiviteit geobjectiveerd moeten worden om zodoende te kunnen meten. De vraag die daaruit voortvloeit is:

2. Hoe kan intuïtiviteit worden vertaald naar meetbare eenheden om zo de verschillen tussen de oorspronkelijke en verbeterde notatie zichtbaar te maken?

Als laatste zal bepaald moeten worden welke experimentele methode kan worden gebruikt voor een vergelijkende vervolgevaluatie van de PGA-modelleertaal.

Hiertoe zal, op basis van de bevindingen uit het literatuuronderzoek, een experimentele methode worden opgezet voor de vervolgevaluatie van de intuïtiviteit van de PGA-modelleertaal. Hiermee kan de derde en laatste deelvraag worden beantwoord:

3. Wat zijn de waargenomen verschillen tussen PGA-2 en PGA-1 in een vergelijkende evaluatie?

De wereldwijde uitbraak van Covid-19 heeft directe invloed op de uitvoering van de experimentele methode. Het verloop van deze uitbraak laat zich niet voorspellen. Om die reden zal gekozen moeten worden voor een vorm waarbij het risico op besmetting met Covid-19 tot nul wordt gereduceerd. De keuze voor een volledig onlineonderzoek is hiermee de meest voor de hand liggende keuze. De experimentele taken zullen dusdanig vorm krijgen dat zij passen binnen een online onderzoeks-omgeving.

1.5. Motivatie/ relevantie

Modelleertalen zoals DSML's vormen in een bedrijfscontext een middel om de bedrijfslogica te begrijpen (Roelens & Poels, 2015)..

Met andere woorden, met een model kan de *Ist*-situatie en/of de *soll*-situatie van een organisatie worden getoond, de werkelijkheid en/of de gewenste werkelijkheid.

Dat toont aan dat het belangrijk is dat een modelleertaal eenvoudig begrepen en toegepast kan worden. Een visuele taal is daarmee een belangrijk middel om die logica over te brengen op verschillende doelgroepen in verschillende (hiërarchische) lagen binnen een organisatie. Zij dragen daarmee in belangrijke mate bij aan de alignment tussen business en IT in de vorm van een gezamenlijk communicatiemiddel

Hoe beter dit middel begrepen kan worden, zonder dat hiervoor voorkennis of training nodig is, hoe beter de alignment tussen business en IT zal zijn. Immers, wanneer een symbool zonder voorkennis of training begrepen kan worden, dan is de kans dat dat symbool universeel is groot. Met een universeel symbool worden misverstanden of verkeerd begrip voorkomen. Dit is een belangrijk punt in bijvoorbeeld het overbrengen van *requirements* voor verbeteringen tussen business en IT(Pollaert, 2016).

Hiermee kan IT uiteindelijk de business optimaal ondersteunen en kunnen kosten door bijvoorbeeld herwerk als gevolg van niet begrijpen worden verminderd of zelfs geëlimineerd. Intuïtiviteit van een EA-modelleertaal draagt daarmee direct bij aan de bedrijfsvoering en de resultaten van organisaties.

Door onderzoek te doen naar de intuïtiviteit van een modelleertaal kan een methode ontwikkeld worden waarmee objectief bepaald kan worden wat de mate van intuïtiviteit is Bork et al. (2019). Hiervan kan in onderzoek naar (door)ontwikkeling van modelleertalen gebruik worden gemaakt om

intuïtiviteit integraal onderdeel van de deze ontwikkeling te laten zijn en draagt dit onderzoek op deze manier bij aan de bestaande literatuur.

1.6. Aanpak in hoofdlijnen

In hoofdstuk 1 is aangegeven wat het onderwerp is van dit onderzoek alsmede de reden om dit onderzoek uit te voeren en de relevantie van dit onderzoek, zowel maatschappelijk als wetenschappelijk. In hoofdstuk 2 zal het theoretisch kader worden onderzocht, besproken en vormgegeven waarin dit onderzoek zijn basis vindt. Vervolgens wordt in hoofdstuk 3 inhoud gegeven aan het ontwerp van het empirisch onderzoek. Hoofdstuk 4 gaat in op de resultaten van het empirisch onderzoek. Ten laatste worden in hoofdstuk 5 de discussie en reflectie, de conclusies en aanbevelingen gedeeld.

2. Theoretisch kader

2.1. Onderzoeksaanpak

De database van de OU-bibliotheek (bibliotheek.ou.nl) is als enige bron gebruikt. Daarnaast zijn alleen artikelen binnen de bedrijfswetenschappen doorzocht. Hiervoor zijn twee redenen: enerzijds de beschikbaarheid van de gevonden literatuur en anderzijds is het een middel om informatieoverflow te voorkomen.

Er is gestart met de Building Blocks methode. De zoek-query's ontstaan door te kijken naar de elementen uit de deelvraag:

*“Welke relevante **methoden** zijn er beschreven om **intuïtiviteit** van **modelleertalen** te **vergelijken** en te **evalueren**?”*

Omdat een **experiment** als onderzoeksmethode eerder is gebruikt en dit per definitie gaat om **empirisch** onderzoek zijn deze woorden toegevoegd aan de lijst met zoektermen. Hiermee wordt beoogd de resultaten zo relevant en zo vergelijkbaar mogelijk te maken met voorgaand onderzoek. Modelleertalen zijn **conceptueel**. Toevoeging van dit woord heeft als doel om mogelijk relevant onderzoek vanuit andere invalshoeken niet uit te sluiten. Van deze elementen is een lijst samengesteld met vertaling, afgeleiden en, voor zover mogelijk, synoniemen (zie tabel 2.1).

Element	Vertaling	Afgeleide	Synoniem/Alternatief
Experimenteel	Experimental	Experiment	
Methodiek	Methodology	Method	
Vergelijkende	Comparing	Compare	
Evaluatie	Evaluating	Evaluation	
Intuïtiviteit	Intuitivity	Intuitive	Understanding
Conceptueel	Conceptual		
Modelleertaal	Modeling Language	Model	Domain Specific Language
Empirie	Empirical		

Tabel 2.1: Overzicht met zoektermen en hun vertaling/synoniem

Voor dit onderzoek is begripsvorming een belangrijke graadmeter voor de (mate van) intuïtiviteit van een modelleertaal. Daarom is **Understanding** meegenomen in het onderzoek. Dit is niet zozeer

gebruikt als zoekterm, maar als indicator van intuïtiviteit tijdens het selecteren van artikelen binnen de zoek-query's.

Als basiswerk is het artikel *"A framework for empirical evaluation of conceptual modeling techniques"* (Gemino & Wand, 2004) genomen. Dit jaartal is als ondergrens voor het publicatiejaar aangehouden. Artikelen van voor deze periode worden voor dit onderzoek als onvoldoende relevant beschouwd. Krantenartikelen en *book reviews* zijn uitgesloten, evenals artikelen waarvan de volledige tekst niet online is te vinden.

Wanneer in de query een door de docent gegeven artikel voorkomt, is dit buiten beschouwing gelaten. Ook artikelen die in meerdere query's voorkomen en die al in een eerdere query zijn geselecteerd, worden geen tweede keer geselecteerd.

2.2. Uitvoering

Op basis van tabel 2.2 is een aantal combinaties van zoektermen opgesteld om de eerste deelvraag te kunnen beantwoorden. Door de zoekquery in te geven in de database van de OU-bibliotheek komen grote aantallen artikelen naar voren, de "hits". Vanwege de beperkte tijd die beschikbaar is voor het onderzoek is de selectie van artikelen beperkt mogelijk. Het stopcriterium dat is gehanteerd is het vijfde artikel dat niet aan de hierna genoemde QuickScan voldeed. De artikelen die gemarkeerd zijn als *"aantal gevonden artikelen"* zijn d.m.v. een *Quick scan* van de abstract van de artikelen. Hierbij is gekeken naar het voorkomen van de zoektermen in de abstract in combinatie met de inhoud daarvan, voor zover hieruit bleek dat het ging om vergelijkende evaluaties in samenhang met modelleertalen of -concepten.

Nr	Query	Aantal hits	Aantal gevonden artikelen	Aantal geselecteerde artikelen	Referenties van relevante artikelen
1	"conceptual model" + "empirical evaluation"	471	4	2	Cherdantseva, Hilton, Rana & Ivins (2016) Bera & Evermann (2014) Bera, Burton-Jones & Wand (2014) Trkman, Mendling, Trkman & Krisper (2019)
2	"Modeling Language"	356	3	1	Reynares, Caliusco & Galli (2014) Burton-Jones & Meso (2006) Roelens & Poels (2015)
3	"Intuitive" + "Modeling Language" + "Experiment"	676	3	2	Farias, Garcia, Wittle, Von Flach Garcia Chavez & Lucena (2015) El-Attar, Luqman, Karpatei, Sindre & Opdahl (2015) Tonella Torchiano, Penta, Ceccato & Ricca (2010)
4	"Domain Specific Modeling Language" + "Intuitive"	43	2	1	Kattenstroth (2012) Bobkowska (2005)

Tabel 2.2: gevonden mogelijk relevante artikelen

Naast de eerdergenoemde selectiecriteria zijn een aantal aanvullende selectiecriteria aangehouden om van de gevonden artikelen de relevantie, op basis van de abstract, zoveel mogelijk te waarborgen.

De criteria om te bepalen of een artikel voldoende relevant is, zijn hieronder benoemd en omschreven.

- **Link met Vergelijken van Concepten:** de betreffende paper vergelijkt concepten van verschillende talen.
- **Link met intuïtiviteit:** in de tekst komt het woord “*intuitive*” of “*understanding*” voor
- **Link met modelleertaal:** er wordt een onderzoek naar modelleertalen gedaan. Dit kan gaan om zowel GPML als om DSML.
- **Link met begripsvorming:** de paper gaat in op of onderzoekt begripsvorming
- **Referentie in Web of Science:** De referentie in *Web of Science* wordt aangehouden als kwaliteitscheck. Het niet hebben van deze referenties leidt niet tot uitsluiting van het artikel, wanneer dit wel aan de overige voorwaarden voldoet.

Tabel 2.3 geeft de waardering per artikel weer. In geval het artikel niet voldoet aan minstens drie van de aanvullende criteria, dan wordt dit artikel als niet relevant voor het onderzoek beschouwd.

Nr	Referentie van relevante artikelen	Link met vergelijken van concepten	Link met Intuïtiviteit	Link met modelleer taal	Link met begrips vorming	Referentie in web of science
1	Cherdantsev, Hilton, Rana & Ivins (2016)	JA	NEE	NEE	JA	-
	Bera & Evermann (2014)	NEE	NEE	JA	JA	9
	Bera, Burton-Jones & Wand (2014)	JA	JA	JA	JA	13
	Trkman, Mending, Trkman & Krisper (2019)	JA	JA	JA	JA	-
2	Reynares, Caliusco & Galli (2014)	JA	NEE	JA	NEE	8
	Burton-Jones & Meso (2006)	NEE	NEE	JA	JA	87
	Roelens & Poels (2006)	JA	JA	JA	JA	4
3	Farias, Carcia, Whittle, von Flach Garcia Chaves & Lucena (2015)	JA	NEE	JA	JA	-
	El-Attar, Luqman, Karpati, Sindre & Opdahl (2015)	NEE	JA	JA	JA	9
	Tonella Torchiano, Penta, Ceccato & Ricca (2010)	NEE	JA	JA	JA	52
	VERVALLEN BIJ ANALYSE					
4	Kattenstroth (2012)	JA	NEE	NEE	NEE	-
	Bobkowska (2005)	JA	JA	JA	JA	-
5	Chacero, Meliá & Hermida (2019) – VERVANGING VOOR VERVALLEN ARTIKEL	JA	JA	JA	JA	-

Tabel 2.3: scores aanvullende criteria per artikel

Tijdens de uiteindelijke analyse van de artikelen bleek een extra query nodig te zijn, omdat vanuit het relevant geachte artikel (Tonella, Torchiano, Penta, Ceccato, & Ricca, 2010) geen (deel)antwoord kon worden geformuleerd op de deelonderzoeksvraag. Deze query staat weergegeven in tabel 2.4

Aanvullende query na analyse gevonden papers						
5	“cognitive Dimensions”+ “modeling language” + “experiment”	3.933	1	1	Cachero, Meliá & Hermida (2019)	

Tabel 2.4: aanvullende zoekquery

2.3. Resultaten en conclusies

2.3.1. Resultaten

Hoewel er, voor zover de onderzoeker bekend, geen direct vergelijkend onderzoek naar intuïtiviteit van modelleertalen lijkt te bestaan, zijn er diverse onderzoeken naar begripsvorming en toepassing van modelleertalen vanuit diverse perspectieven gedaan. In onderstaande tabel 2.5 staan, de naam en auteur(s) van de geanalyseerde artikelen weergegeven, alsmede het doel van het betreffende onderzoek.

In zijn algemeenheid geldt dat de artikelen verschillende benamingen gebruiken om de begrippen “modelleertaal” of “notatie” aan te duiden. In dit rapport worden de verschillende termen benoemd met “modelleertaal” en “notatie”.

Waar de meeste artikelen tot doel hebben om het effect van een of meerdere visuele notaties te onderzoeken op begripsvorming, kijkt het onderzoek van Trkman, Mendling, Trkman, and Krisper (2019) specifiek naar het verschil in impact van een tweetal verschillende business proces formats, tekstuele *use cases* en visuele BPMN-modellen, op het domeinbegrip. Hieruit blijkt o.a. dat visuele notaties tot een beter resultaat leiden in probleemoplossende taken. Dit wordt ondersteund door het onderzoek van (Cachero, Meliá, & Hermida, 2019) die een vergelijkbaar onderzoek hebben uitgevoerd onder junior *software developers* en het effect van twee verschillende *business proces formats* op domeinmodelcompositie. Hieruit volgde dat deze beter in staat waren een domeinmodel te construeren, wanneer zij gevoed werden door visuele UML-notatie met informatie over het betreffende domein dan wanneer dat met tekstuele notaties gebeurde.

Deze onderzoeken geven hiermee aan dat visuele talen tot een betere begripsvorming bij probleemoplossende taken leiden. Dit suggereert een hogere mate van intuïtief zijn bij visuele notaties dan bij tekstuele.

Artikel	Doel
Bera, Burton- Jones & Wand (2014)	<ol style="list-style-type: none">1. Testen of ontologische duidelijkheid van de modelleertaal helpt het domein beter te begrijpen2. Testen of domein(voor)kennis van proefpersonen of het ontbreken daarvan de voordelen van de onthologische duidelijkheid elimineert
Trkman, Meding, Trkman & Krisper (2019)	Het vergelijken van twee verschillende <i>business procesformats</i> : tekstuele <i>usecases</i> en visuele BPMN-modellen
Roelens & Poels (2015)	Ontwikkelen en evalueren van een nieuw <i>businessmodel viewpoint</i> dat helpt een beter begrip te krijgen van de onderliggende kennis van de business
El-Attar, Luqman, Karpati, Sindre & Opdahl (2015)	Het effect vergelijken van het gebruik van nieuwe en originele notaties om veiligheidsaspecten in situ te modelleren. Evaluatie is voornamelijk gebaseerd op de modelleringsmogelijkheden van twee notaties in plaats van op de perceptie van de modelgebruikers Uitindelijke doel van dit onderzoek is om praktische beveiligingsmodellering te verbeteren
Farias, Garcia, Whittle, von Flach Garcia Chavez & Lucena (2015)	De impact evalueren van modelcompositie technieken op de inspanning van ontwikkelaars en model correctheid bij het ontwikkelen van ontwerpmodellen
Bobkowska (2005)	De toepassing van een CD-vragenlijst als methode voor VML-evaluatie evalueren door deze in de praktijk te proberen.
Cachero, Meliá & Hermida (2019)	Het effect (efficiëntie en effectiviteit) van het gebruik van een tekstuele vs. het gebruik van een grafische notatie voor de representatie van domeinmodellen

	te beoordelen vanuit het oogpunt van junior softwareontwikkelaars die individueel een softwaresysteem modelleren.
--	---

Tabel 2.5: geanalyseerde artikelen

De verschillende hypothesen en bijbehorende maatstaven zijn weergegeven in tabel 2.6. Deze hypothesen hebben vrijwel allemaal met elkaar gemeen dat zij gaan over de begripsvorming van een notatie of modelleertaal. Uitzondering hierop vormt het onderzoek van (Bobkowska, 2005), dat de evaluatie van een evaluatiemethode van modelleertalen betreft en niet de evaluatie van modelleertalen zelf.

Het aantal juiste antwoorden en de benodigde tijd hiervoor, komen in alle onderzoeken terug als maatstaf. Deze maatstaven zijn daarmee zeer relevant wanneer het gaat om het objectief kunnen meten van begrip en toepassing van de notatie. De juiste antwoorden geven de effectiviteit van het begrip weer, de benodigde tijd geeft de efficiëntie van het begrip weer. Wanneer deze aan elkaar gerelateerd worden, wordt hiermee de doeltreffendheid weergegeven.

Roelens and Poels (2015) spreken meer specifiek over “Interpretatieve effectiviteit”, de nauwkeurigheid waarmee de modelleertaal begrepen wordt en informatie hieruit gedestilleerd wordt, gemeten in juiste antwoorden en “Interpretatieve efficiëntie”, de snelheid waarmee de notatie wordt begrepen, gemeten in tijd. Zij baseren zich hierbij op Gemino and Wand (2004).

<i>Hypothesis</i>	<i>Measures</i>	<i>Experimental Design</i>
Art 1: Het voordeel van onthologische duidelijkheid zal groter zijn bij een gematigde domeinkennis conditie dan bij H1 De lage domeinkennis conditie (H1) H2 De hoge domeinkennis conditie	performance van testpersoon	<i>Mixdesign within/between</i>
Art 2 H1. Identificatie van userstories is nauwkeuriger wanneer dit wordt ondersteund door een tekstueel gebruiksmodel dan door een BPMN-model. H2. Het begrip van de uitvoeringsvolgorde-afhankelijkheden van een userstory is groter wanneer ondersteund door een BPMN-model dan door een tekstueel gebruiksmodel. H3. Het begrip van de integratieafhankelijkheden van een userstory is groter wanneer het wordt ondersteund door een BPMN-model dan door een tekstueel gebruiksmodel.	performance van testpersoon	<i>Mixdesign within/between</i>
Art 3 Hc: de begripseffectiviteit (Hc1), efficiëntie (Hc2) en de werkzaamheid (Hc3) van het nieuwe BM-gezichtspunt is hoger dan de Hc1, Hc2 en Hc3 van de bestaande VDML-diagrammen.	Interpretatieve Effectiviteit: % juiste antwoorden op begripsvragen Absolute aantallen juiste antwoorden op probleemoplossende vragen Tijd als maatstaf voor Interpretatieve Efficiëntie Verhouding tussen % juiste antwoorden en absolute aantallen	<i>Mixdesign within/between</i>

<p>Art 4 (Ho1): De responsetijd van de uitgebreide notatie \geq Ha1): De responsetijd van de uitgebreide notatie $<$ de responsetijd van de oorspronkelijke notatie((Ho2): De fouten door de uitgebreide notatie \geq de fouten door de oorspronkelijke notatie (Ha2): De fouten door de uitgebreide notatie $<$ de fouten door de oorspronkelijke notatie</p>	<p>H 1: Responsetijd H 2: fouten</p>	<p><i>Mixdesign within/between</i></p>
<p>Art 5 Nulhypothese H1: De op specificatie gebaseerde compositietechniek vereist minder (of gelijke) inspanning dan de op heuristisch gebaseerde compositietechniek om ModelAB te produceren van ModelA en ModelB. Alternatieve hypothese H1: De op specificatie gebaseerde compositietechniek vereist meer inspanning dan de op heuristische gebaseerde compositietechniek om MAB te produceren van MA en MB. Nulhypothese H2: De op specificatie gebaseerde compositietechniek levert een lager (of gelijk) aantal correct samengestelde modellen op dan de op heuristische gebaseerde compositietechniek. Alternatieve hypothese H2: De op specificatie gebaseerde compositietechniek levert een groter aantal correct samengestelde modellen op dan de op heuristisch gebaseerde compositietechniek.</p>	<p>H1: Inspanning -> tijd in minuten H2: Juistheid en inconsistentie percentage</p>	<p><i>Within subjects</i></p>
<p>Art 6 meer gedetailleerde vragenlijsten helpen bij het verkrijgen van meer gedetailleerde gegevens.</p>	<p>Gebruiksgemak Flexibiliteit Effectiviteit Efficiëntie</p>	<p><i>Within subjects</i></p>
<p>Art 7 • HProductivity0: Het gebruik van de grafische of tekstuele notatie van OOH4RIA voor domeinmodellering heeft geen invloed op de productiviteit van de softwareontwikkelaars (d.w.z. het</p>	<p>Percentage correct geïdentificeerde elementen Ontwikkeltijd</p>	<p><i>Mixdesign within/between</i></p>

percentage correct geïdentificeerde elementen en de ontwikkeltijd), onafhankelijk van het systeem dat wordt ontwikkeld. • HEffectivenessAttr0: Het gebruik van de grafische of tekstuele notatie van de benadering OOH4RIA voor domeinmodellering heeft geen invloed op het percentage attributen dat correct wordt gedetecteerd, onafhankelijk van het systeem dat wordt ontwikkeld.		
--	--	--

Tabel 2.6 : hypothesen en maatstaven

In de onderzoeken gaat het, zoals tabel 2.6 laat zien, zonder uitzondering om een experimentele onderzoeksmethode waar grotendeels is gekozen voor een *experimental design* in een mix van *within subjects* and *between subjects*. Een *within subjects* design betekent dat een groep respondenten aan dezelfde condities worden blootgesteld. Een voordeel hiervan is dat individuele verschillen tussen respondenten weinig tot geen invloed hebben op het te onderzoeken verband. Een groot nadeel van deze methode bestaat hieruit dat zich de situatie voor kan doen dat bepaalde condities elkaar versterken en daarmee de resultaten beïnvloeden.

Een *between subjects* design, ook wel *between groups* genoemd, verdeelt de respondenten in twee groepen, waarbij een groep de controlegroep vormt. Deze controlegroep krijgt niet dezelfde condities als de onderzoeksgroep. Dit ontwerp houdt geen rekening met individuele verschillen, wat een nadeel vormt van deze methode.

Door een mix te creëren van beide ontwerpen, kunnen de nadelen van de methoden worden gemitigeerd door het gebruik van een testgroep en een controlegroep en het verschil in behandeling tussen beide groepen. Hierbij wordt de testgroep als eerste blootgesteld aan de nieuwe of gemanipuleerde behandeling en de controlegroep aan de oorspronkelijke behandeling. In de volgende taak wordt deze behandeling omgedraaid.

Bij de praktische uitvoering (zie tabel 2.7), vindt in een aantal gevallen een trainingsfase plaats om de respondenten kennis te laten maken met de betreffende modelleertaal of notatie (Bera, Burton-Jones, & Wand, 2014) (Farias, Garcia, Whittle, von Flach Garcia Chavez, & Lucena, 2015).

Verder bestaan de onderzoeken uit opdrachten die veelal m.b.v. een combinatie van probleem-oplossende vragen en begripsvragen moeten worden opgelost door de respondenten. In een aantal gevallen is naast de opdrachten ook een vragenlijst toegevoegd die als doel heeft het ervaren begrips-en/of gebruiksgemak te meten vanuit de perceptie van de respondent. Het onderzoek van (Bobkowska, 2005) zegt hierover dat vanwege de vele parameters die moeten worden meegenomen, het evalueren van een visuele modelleertaal welhaast onbegonnen werk is, waarbij het reflecteren van respondent op resultaten van uitgevoerde taken de enige manier lijkt te zijn om te kunnen evalueren. Desondanks vindt zij dat een theoretische analyse of een beproefde methode als basis voor een evaluatie niet zonder meer kan worden verworpen.

Wanneer het gaat om keuze voor respondenten wordt er iets meer gebruik gemaakt van studenten en iets minder van professionals. Voor een groot deel heeft dit te maken met praktische overwegingen (e.g. beschikbaarheid). Daarnaast valt op dat er ofwel geen tijdslimiet is aangehouden binnen het onderzoek, ofwel er is binnen het onderzoeks-ontwerp rekening gehouden met de maximaal beschikbare tijd in bijvoorbeeld een klassessie. Verder is in alle onderzoeken, waar sprake is van een

gemixt ontwerp, toewijzing aan de test-of controlegroep willekeurig en is anonieme verwerking van de resultaten gewaarborgd. In een aantal gevallen vroeg de aard van het onderzoek een trainingssituatie m.b.t. de gebruikte notaties of modellen. Daarnaast is met de onderzoeks-specifieke omstandigheden rekening gehouden, zoals bijvoorbeeld de beschikbare tijd als vast gegeven binnen een klassessie, of het aanpassen van de complexiteit als gevolg de beschikbare tijd als vast gegeven. Deze elementen verschillen echter per onderzoek en laten geen gemene deler zien.

<i>Instrumentation & Experimental tasks</i>	<i>Participants</i>	<i>Operational Procedures</i>
Art 1 - modelleringsscript lezen - antwoorden te geven op inferentiële probleemoplossende vragen over het domein dat in de modelleertaal wordt getoond.	Studenten	1. training fase bestaande uit - enquête betreffende modelleerkennis en domein kennis - 15 minuten reviewtijd om basis <i>extended entity relationship</i> concepten te bestuderen - oefening beantwoorden probleemoplossende vragen binnen een eenvoudige casus 2. <i>Main study</i> bestaande uit twee taken. - het beschrijven van de inhoud van de modelleertaal zoals zij dit begrepen om de respondent kennis te laten maken met de modelleertalen - hoofdtak: beantwoorden van probleemoplossende vragen over een domein - 2 x herhalen van de hoofdtak met andere domeinen
Art 2 Begripstest (controle doel) Taak 1: Identificeren <i>userstories</i> (H1) Taak 2: uitvoeringsvolgorde afhankelijkheden begrijpen (H2) Taak 3: Integratieafhankelijkheden begrijpen (H3) <i>Recall</i> test (controle doel) Vragen over ervaren bruikbaarheid (informatief doel)	Studenten	- 3 pilot-experimenten om begrip van het materiaal te testen - werkboeken dubbelzijdig geprint in landscape - alle experimenten volgden hetzelfde protocol: introductie van 2 onderzoekers. Een onderzoeker leidde het experiment, de andere onderzoeker bood ondersteuning en hulp - duur: tussen 67 en 74 minuten

<p>Art 3</p> <p>vier sets diagrammen: de bestaande VDML-gezichtspunten en het nieuwe BM-gezichtspunt toegepast op de gezondheidszorg en de productiecase.</p> <p>De experimentele taken omvatten voor beide gevallen dezelfde begripsvragen en probleemoplossende vragen</p> <p>De experimentele taken zijn vooraf getest om de formulering van de instructies en de vragen te verifiëren.</p>	<p>Studenten</p>	<ul style="list-style-type: none"> - geïmplementeerd als een vrijwillige oefening in de klas. - willekeurige toewijzing aan vier verschillende slots - geïnformeerd dat de antwoorden anoniem worden verwerkt, het experiment op elk moment kan worden afgebroken en de taken in eigen tempo kunnen worden uitgevoerd. - vragen door de ene onderzoeker opgelost en door de andere gevalideerd. - één punt toegewezen voor elk correct antwoord, terwijl een half punt wordt afgeleid voor aanvullende antwoorden - De antwoorden op de probleemoplossende vragen worden gecorrigeerd door drie onderzoekers die onderscheid maken tussen goede en foute antwoorden. - De eindscore wordt behaald door één punt toe te kennen aan de antwoorden, die door alle onderzoekers als correct worden beschouwd.
<p>Art 4</p> <p>1. tutorial over de notaties om de proefpersonen vertrouwd te maken met de notaties en hun onderliggende semantiek.</p> <p>2. Sessie met ongelimiteerde tijd waarin de proefpersonen de toegewezen experimentele taken uitvoerden.</p>	<p>Professionals</p>	<ul style="list-style-type: none"> - Alle experimentele artefacten werden in kleur op papier gedrukt. - Zwart-witmodellen werden ook afgedrukt op papier van hetzelfde formaat. - De proefpersonen werd verzocht om vragen op papier te beantwoorden met potloden, die aan alle proefpersonen werden verstrekt. - De proefpersonen gebruikten ook stopwatches die beschikbaar waren op hun eigen mobiele telefoons om de tijd bij te houden. - De proefpersonen zijn niet geïnformeerd over de te onderzoeken hypothesen

<p>Art 5</p> <ul style="list-style-type: none"> - Training: om ervoor te zorgen dat de proefpersonen de nodige bekendheid verwerven met elke modelcompositietechniek. - Pas de technieken: stel MA en MB samen op basis van een beschrijving van veranderingen die definieert hoe de modelelementen van MA werden veranderd. Elk proefpersoon heeft deze taak zes keer uitgevoerd. - Detecteer inconsistenties Proefpersonen beoordeelden MCM om inconsistenties te detecteren. - Los inconsistenties op De proefpersonen hebben de inconsistenties opgelost die eerder waren gelokaliseerd om MAB te produceren. - Semi gestructureerd interview waarin respondenten middels een vragenlijst reflecteren 	<p>Professionals en studenten</p>	<p>De proefpersonen gebruikten drie compositietechnieken (d.w.z. Epsilon, IBM RSA en traditionele algoritmen) om zes evolutiescenario's uit te voeren en waren noch bekend met dergelijke scenario's, noch met de gebruikte ontwerpmodellen.</p> <p>De gebruikte ontwerpmodellen waren fragmenten van industriële modellen die werden vastgelegd uit verschillende toepassingsdomeinen, waaronder financiële en simulatie van oliewinning.</p>
<p>Art 6</p> <p>Twee fasen met twee taken</p> <p>Fase 1</p> <ul style="list-style-type: none"> - UML en de geselecteerde CASE-tool evalueren door de CD-vragenlijst in te vullen; - reflecteren en commentaar geven op de resultaten die zijn bereikt met de methode en meningen uiten over het nut ervan voor evaluatie. <p>Fase 2</p> <ul style="list-style-type: none"> - Use Cases evalueren door de CD-VML-UC-vragenlijst in te vullen; - reflecteren en commentaar geven op de resultaten die zijn bereikt met de methode en meningen uiten over het nut ervan voor evaluatie. 	<p>Professionals en studenten</p>	<p>Fase 1</p> <ul style="list-style-type: none"> - CD-vragenlijst is negen pagina's lang - studie duurde ongeveer een uur - ingevuld en geanalyseerd aan de hand van criteria voor een goede methodologie. <p>Fase 2</p> <ul style="list-style-type: none"> - CD-VML-UC-vragenlijst was achttien pagina's lang - studie duurde ongeveer vier uur. - Voorafgaand aan het invullen van de vragenlijst bij onderwerpen een tutorial met informatie over de aanpak en de rol van studenten in de studie, de theorie van VML, informatie over cognitieve dimensies en structuur van de vragenlijst. - Tijdens de studie kon UML standaarddocumentatie worden gebruikt
<p>Art 7</p> <p>Stel een domeinmodel op, op basis van de gegeven tekstuele of visuele notaties betreffende het domein</p>	<p>Studenten</p>	<ul style="list-style-type: none"> - Ontworpen als klassessie - duur: 2 uur - complexiteit aangepast aan tijdslimiet - Willekeurige toewijzing aan 2 factoren in 2 varianten: notatie (grafisch of tekstueel) en systeem (ticketverkoper of hotelmanager). - Beide typen notatie hebben dezelfde abstracte syntaxis en semantiek, maar verschillen in hun concrete syntaxis of notatie.

Tabel 2.7: praktische uitvoering

2.3.2. Conclusie

In hoofdstuk 1 zijn twee deelvragen gesteld waarvan de antwoorden moeten gaan leiden tot beantwoording van de centrale onderzoeksvraag.

1. Welke relevante methoden zijn er beschreven om modelleertalen te vergelijken en te evalueren?

Met betrekking tot vraag 1 kan worden gesteld dat er in de literatuur voldoende verwijzingen zijn te vinden naar gebruikte methoden om modelleertalen te evalueren en te vergelijken. Binnen de geraadpleegde literatuur gaat dit zonder uitzondering om een experiment, waarbij de voorkeur lijkt uit te gaan naar een gemixt experimenteel ontwerp. Hierbij moet een mix van begrips- en probleemoplossende vragen bijdragen aan de antwoorden op de vragen die binnen de onderzoeken worden gesteld. Daarnaast zijn herinneringstechnieken een middel om begripsvorming te toetsen.

Een visuele taal kent haar afhankelijkheid van menselijke creativiteit. Uit onderzoek is ook gebleken dat de reflectie van respondenten op het ervaren gebruiksgemak van de notatie een belangrijke rol speelt.

Zoals aangegeven gaat het in alle onderzoeken in meer of mindere mate om begripsvorming en is begripsvorming een belangrijk element wanneer het gaat om het intuïtief kunnen begrijpen en toepassen van de notaties binnen een modelleertaal.

In de eerste plaats blijkt uit het onderzoek dat het visuele aspect leidt tot het sneller begrijpen **en** kunnen toepassen van een notatie binnen experimentele taken. Omdat de PGA-taal bestaat uit visuele notaties kan op grond van eerder onderzoek worden gesteld dat de notaties van de PGA-taal in beginsel al meer intuïtief zijn.

Uit de literatuur blijkt dat er drie indicatoren zijn om intuïtiviteit te meten. Hierbij gaat het om:

1. Juistheid van de gegeven antwoorden binnen experimentele taken
2. De snelheid waarmee de respondenten tot een (juist) antwoord komen
3. De mate waarin respondenten een notatie als intuïtief ervaren

In hoofdstuk 3 wordt ingegaan op de specifieke methode die kan worden gebruikt met betrekking tot dit onderzoek.

2.4. Doel van het vervolgonderzoek

De onderzoeker wil de vraag beantwoorden hoe de intuïtiviteit van EA-modelleertalen vergelijkend kan worden geëvalueerd om zodoende vast te kunnen stellen of de verbetering zich ook daadwerkelijk heeft voorgedaan ten opzichte van de oorspronkelijke PGA-notatie. Om deze vraag te kunnen beantwoorden zal intuïtiviteit een concrete waarde moeten krijgen. Door het meetbaar maken van intuïtiviteit kunnen (versies van) talen met elkaar worden vergeleken.

Daarnaast zal een experimentele methode uitgevoerd worden waarmee de intuïtiviteit van EA-Modelleertalen kan worden gemeten en vergeleken. Deze methode zal toegepast worden op de PGA-1 notatie en de PGA-2 notatie. Op deze wijze kan worden bepaald of de laatstgenoemde, verbeterde, PGA-notatie significant beter presteert.

Dit geeft enerzijds antwoord op de vraag in hoeverre de intuïtiviteit van de PGA-2 notaties is verhoogd. In breder verband geeft dit onderzoek een methodiek om de intuïtiviteit van modelleertalen te meten en te vergelijken.

3. Methodologie

3.1. Conceptueel ontwerp: keuze van onderzoeksmethode(n)

Dit deel van het onderzoek heeft tot doel het begrip “Intuïtiviteit van de notatie” meetbaar te maken en te objectiveren om zodoende twee versies van dezelfde modelleertaal (PGA-1 en PGA-2) met elkaar te kunnen vergelijken en daaruit conclusies te kunnen trekken.

De experimentele techniek van Bork et al. (2019) is gericht op het ontwikkelen van een meer intuïtieve notatie. Hiermee is echter niet te bepalen of de verbetering zich ook daadwerkelijk voordoet. Daarnaast is de analyse en ontwikkeling van de nieuwe notatie deels afhankelijk van menselijke creativiteit en daarmee lastig te objectiveren en generaliseren.

Om objectief te kunnen vaststellen of de verbetering zich ook daadwerkelijk heeft voorgedaan is het volgende nodig:

1. Relevante methoden om intuïtiviteit van notaties vergelijkend te evalueren. Deze zijn beschreven in hoofdstuk 2
2. Een gedetailleerde onderzoeksmethode om de intuïtiviteit van de PGA-modelleertaal vergelijkend te evalueren. Deze wordt beschreven in dit hoofdstuk, hoofdstuk 3.
3. Vergelijkende evaluatieresultaten van de intuïtiviteit van de PGA-modelleertaal in een praktische bedrijfscontext. Deze zullen worden besproken in hoofdstuk 4.

Deze punten kunnen worden gevonden in de methodologische literatuur enerzijds en de praktische bedrijfscontext anderzijds.

In dit onderzoek moet gezocht worden naar een methode om een subjectief begrip als intuïtiviteit objectief meetbaar te maken. Hierbij wordt de definitie van intuïtiviteit, het zonder voorkennis kunnen begrijpen van dingen, van Jošt et al. (2016) aangehouden. Met andere woorden, een methode om begripsvorming en of begrijpelijkheid te meten is voor dit onderzoek de relevante methode om data te verzamelen.

Omdat het gaat om een onderzoek waarbij objectivering van evaluatieresultaten een grote rol speelt, wordt gekozen voor een deductieve aanpak, waarbij vanuit de theorie hypothesen worden geformuleerd en deze hypothesen door middel van dataverzameling worden getoetst. Door het onderzoek te stelen op de theorie wordt een zekere mate van objectiviteit verworven. Uit de theorie is gebleken dat een experiment met een gemengd experimenteel ontwerp de meest gebruikte onderzoeksmethode is om begripsvorming en begrijpelijkheid te onderzoeken. Een gemengd ontwerp is een type “*Multiple methods research*” dat het gebruik van kwalitatieve en kwantitatieve dataverzamelingstechnieken en analytische procedures in hetzelfde onderzoek integreert” (Saunders, 2019). Deze verschillende technieken zijn verweven in de taken die de respondenten opgegeven krijgen en die worden beschreven in de volgende paragrafen.

3.2. Technisch ontwerp: uitwerking van de methode

3.2.1. Variabelen en maatstaven

Uit de literatuur is gebleken dat effectiviteit en efficiëntie van taakuitvoering goede maatstaven zijn om intuïtiviteit te meten.

Roelens and Poels (2015) beschrijven de relatie tussen efficiëntie en effectiviteit enerzijds en begripsvorming anderzijds door te spreken over Interpretatieve efficiëntie, gemeten door de

benodigde tijd in minuten, en Interpretatieve effectiviteit gemeten door het percentage juiste antwoorden. Deze maatstaven vormen een objectieve manier om de afhankelijke variabelen interpretatieve efficiëntie en interpretatieve effectiviteit te kunnen meten.

Uit het onderzoek van (Bobkowska, 2005) blijkt dat een persoonlijke reflectie m.b.t hoe de taal ervaren wordt een belangrijk element is dat tot evaluatie van een visuele notatie hoort. Deze persoonlijke reflectie is een subjectieve manier om de ervaren intuïtiviteit als afhankelijke variabele te meten.

Binnen de evaluatie die in dit onderzoek wordt uitgevoerd worden de afhankelijke variabelen gevormd door enerzijds de oorspronkelijke PGA notatie (PGA-1) en anderzijds de nieuwe PGA notatie (PGA-2).

3.2.2. Onderzoeksontwerp

Om te toetsen of de PGA-2 notatie ook daadwerkelijk beter presteert wordt een vergelijkende evaluatie uitgevoerd. Dit betreft een experiment waarop een gemengd ontwerp wordt toegepast.

Gezien de beperkte beschikbare tijd wordt gekozen voor een relatief kleine onderzoeksgroep. Deze wordt verdeeld in een A en B groep op basis van geboortjaar (even - vs. oneven geboortjaar). Per element krijgt een respondent ofwel PGA-1, ofwel PGA-2 te zien. Dit zorgt ervoor dat de respondenten niet weten tot welke versie (PGA-1 of PGA-2) een getoond symbool behoort.

In de eerste drie taken wordt de interpretatieve effectiviteit (juiste antwoorden) en de interpretatieve efficiëntie gemeten. Dit zijn objectieve meeteenheden die binnen een *between subjects design* zullen worden gemeten.

Deze werkwijze geldt voor beide groepen. Waar groep A voor een element de PGA-1 notatie te zien krijgt, krijgt groep B voor hetzelfde element de PGA-2 notatie te zien. Hiermee zijn beide groepen op dezelfde manier voorbereid op de laatste taak, waarmee de ervaren intuïtiviteit wordt gemeten.

Met betrekking tot de ervaren intuïtiviteit geldt dat er sprake is van een subjectieve meeteenheid, immers, het gaat om een persoonlijke mening van een respondent. Dit wordt binnen een *within-subject design* gemeten.

De keuze voor een gemengd ontwerp van het experiment is eveneens gerechtvaardigd in de beperkte schaal van de experimentele groep.

3.2.3. Experimentele taken en wijze van toetsing

Zoals gebleken is uit het onderzoek van (Bera et al., 2014) hebben respondenten met een matige domeinkennis het meest profijt van een intuïtieve notatie. Mede om die reden worden zowel het type casus als de volgorde waarin de respondenten worden blootgesteld aan PGA-1 en PGA-2 opgesteld als een *between subjects* ontwerp voor de eerste drie experimentele taken en een *within-subjects* ontwerp voor de laatste experimentele taak. Dit heeft als doel om eventueel aanwezige voorkennis te mitigeren en een lerend effect te voorkomen.

Het onderzoek start met demografische vragen. Deze zijn opgenomen in onderstaande tabel 3.1. Deze vragen hebben enerzijds als doel de populatie te kunnen indelen in een van beide onderzoeksgroepen op basis van het geboortjaar, anderzijds is hiervan het doel om eventuele effecten als gevolg van bijvoorbeeld modelleerervaring te kunnen verklaren.

Demografische vragen
Geslacht
Geboortedatum
Opleidingsniveau (hoogst genoten, afgerond met diploma)
Werksector (publiek, privaat, academisch)
Modelleerervaring (in jaren)
Modelleerexpertise (in talen)
Aantal modellen gezien
Aantal modellen gemaakt

Tabel 3.1: Demografische vragen

Vervolgens wordt voorafgaand aan de experimentele taken wordt een summiere uitleg over de PGA-modelleertaal getoond. Deze uitleg is beperkt tot het doel van de taal en hoe deze taal gebruikt wordt. Deze uitleg wordt weergegeven in afbeelding 3.1.



PGA modeling technique is oriented toward the realization of strategic fit within the business architecture.

What?

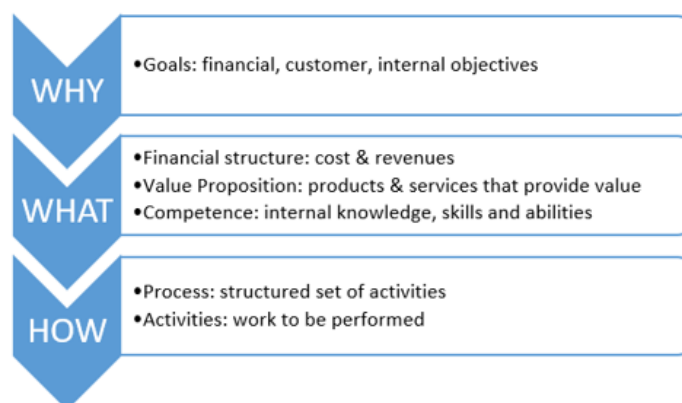
Strategic fit in the business architecture:
translating strategies into concrete activities and results

65% of organizations have an agreed-upon strategy

14% of the employees understand this strategy

10% of organizations successfully **execute** the strategy

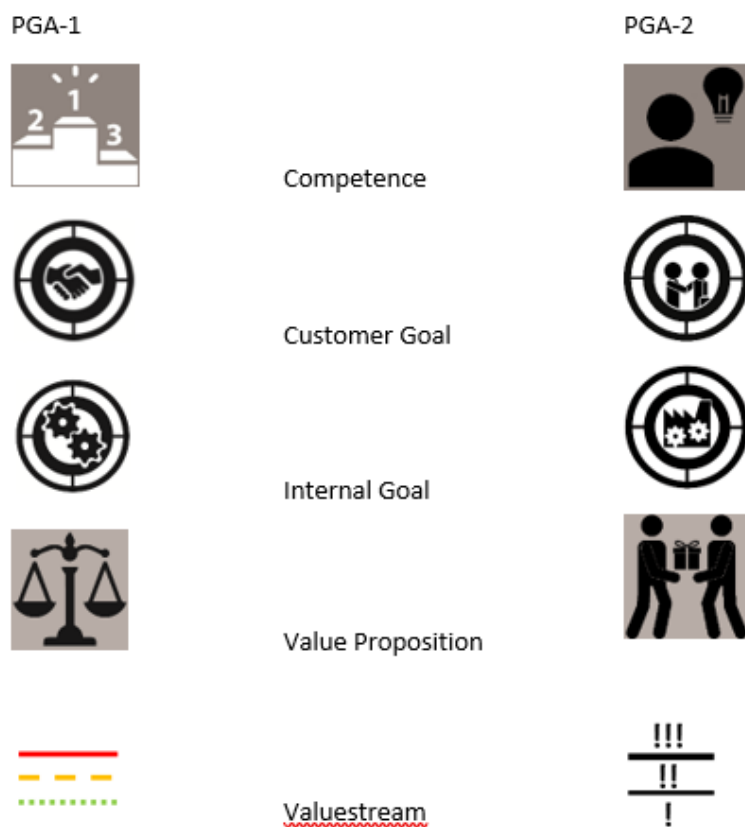
How?



Afbeelding 3.1: uitleg PGA-taal

Om de hypothesen te kunnen toetsen worden, zoals aangegeven, de respondenten een viertal experimentele taken voorgelegd waarin zij zullen worden blootgesteld aan een mix van symbolen uit PGA-1 en PGA-2.

Hierbij is uitsluitend gebruik gemaakt van de symbolen die in PGA-2 zijn veranderd ten opzichte van PGA-1. Het gaat om de symbolen die zijn weergegeven in onderstaande afbeelding 3.2



Afbeelding 3.2: uitgevraagde symbolen PGA-1 en PGA-2

De effectiviteit van de taal kan worden gemeten met begripsvragen, probleemoplossende vragen en herinneringstechnieken (Gemino & Wand, 2004). In dit onderzoek gaat het om de intuïtiviteit van de taal. Om die reden is gekozen voor een combinatie van begripsvragen en herinneringstechnieken. Hierbij speelt domeinbegrip een rol, maar vooral gaat het om begrip van de symbolen van de taal zelf.

Met de experimentele taken een tot en met drie kan worden gemeten of er verschil bestaat tussen de twee groepen wanneer het gaat om snelheid en juistheid van antwoorden. Hiermee worden de afhankelijke variabelen interpretatieve effectiviteit en interpretatieve efficiëntie gemeten.

Taak 1: Notatie-associatie

Elke respondent wordt geconfronteerd met een mix van symbolen uit PGA-1- resp. PGA-2 , waarbij de respondenten tot drie associaties kunnen geven. Beantwoording van de vraag is verplicht. Zoals aangegeven worden de respondenten per symbool blootgesteld aan ofwel PGA-1, ofwel PGA-2.

De scoring wordt bepaald door de juistheid van de antwoorden, waarmee de afhankelijke variabele: interpretatieve effectiviteit wordt gemeten. Daarnaast wordt de tijd, benodigd voor het beantwoorden van de vraag, gemeten. Hiermee wordt de interpretatieve efficiëntie bepaald.

De volgorde waaraan beide groepen zijn blootgesteld is opgenomen in bijlage 1: Notatie Associatie. Hierin is ook de scoring en het antwoordmodel opgenomen. Het controlemodel zal

bestaan uit een lijst met juiste antwoorden. De mogelijke scores worden onderverdeeld in 4 categorieën, zijnde juist, gedeeltelijk juist en onjuist. De vierde categorie bestaat uit een foute associatie van het getoonde symbool met een ander element uit de PGA-taal. Het associëren van het getoonde element met een ander symbool uit de PGA-taal is in het bijzonder ongewenst. Dit wijst namelijk op een begripsverwarring, waarmee intuïtiviteit wordt ondermijnd.

Door te kiezen voor een testmethode waarbij “de eerste ingeving” wordt gestimuleerd wordt getracht zo dicht mogelijk bij een intuïtieve associatie te komen. Deze taak wordt voor het eerst gebruikt door Bork et al. (2019), specifiek om intuïtiviteit van een notatie te meten. Dit sluit aan bij de door Jošt et al. (2016) gegeven definitie van intuïtiviteit in relatie tot modelleertalen; een notatie die zonder voorkennis of training wordt begrepen.

Taak 2: Begrip – Comprehension Questions

Het experiment-domeinmodel vanuit een mix van PGA-1 en PGA-2 wordt beschikbaar gesteld aan de respondenten. Zij krijgen daarbij meerkeuzevragen die de kennis over het voorgestelde domein, een service voor het online bestellen van maaltijden, toetsen. Deze zijn, samen met het scoringsmodel, opgenomen in bijlage 2.

De keuze voor de bedrijfscontext is ingegeven door de vanuit de literatuur gebleken vooronderstelling dat een visuele notatie het meeste resultaat biedt wanneer de respondenten een beperkte voorkennis hebben van de betreffende context.

Taak 3: Herinneren – Recall Questions

In de derde taak worden de respondenten gedurende 1 minuut blootgesteld aan de symbolen met hun betekenis en gevraagd deze in te prenten. Vervolgens worden zij, per element en zonder de betekenis, opnieuw blootgesteld aan de symbolen en gevraagd hierbij de juiste betekenis te geven.

Naast de antwoorden van de respondenten is ook de besteedde tijd per vraag voor alle drie de taken gemeten. Hierbij moet worden aangetekend dat zij gedurende alle eerdere taken steeds met dezelfde variant van de PGA-taal werden geconfronteerd per symbool.

Voor zowel de notatie-associatie als de begrips- en herinneringsvragen geldt dat wanneer een respondent een symbool binnen een versie van PGA heeft gezien, deze versie voor het betreffende symbool in de daaropvolgende taken wordt aangehouden.

Naast de interpretatieve effectiviteit en de interpretatieve efficiëntie wordt ook de ervaren intuïtiviteit gemeten. Dit gebeurt met de vierde en laatste experimentele taak.

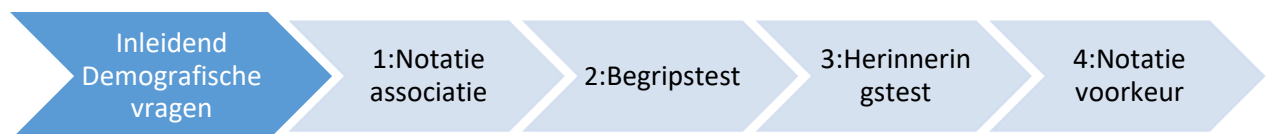
Taak 4: Notatie voorkeur

De respondenten krijgen beide notaties te zien. Hierbij wordt hen gevraagd aan te geven welke notatie zij als meer intuïtief beschouwden en daarbij hun antwoord te motiveren

Zoals uit het onderzoek van Bobkowska (2005) is gebleken, blijft de persoonlijke reflectie een belangrijk element om de intuïtiviteit te bepalen. In de laatste taak wordt daarom de respondenten gevraagd een notatievoorkeur aan te geven. Hierbij krijgen zij beide varianten van de notatie te zien, waarbij de variant van de notatie waaraan zij zijn blootgesteld zich steeds aan de linkerkant van de schaal vindt en het symbool van de andere variant aan de rechterkant van de schaal. Door een positie op de schaal te verplaatsen in de richting van het symbool wat de respondent als meer intuïtief ervaart kan de notatievoorkeur worden aangegeven. Daarnaast zal de respondent worden gevraagd een motivatie voor zijn of haar keuze te geven als reflectie. Deze motivatie is facultatief.

Met al deze taken tezamen kan een vergelijkende evaluatie worden uitgevoerd van de PGA-2 taal ten opzichte van de PGA-1 taal.

Het onderzoeksproces ziet er hiermee als volgt uit:



3.2.4. Hypothesen

In de vergelijking van de oorspronkelijke PGA-1 en de verbeterde PGA-2 notatie bestaat de verwachting dat de PGA-2 een betere score zal laten zien. De verbeterde notatie is immers tot stand gekomen met de creatieve inbreng van testpersonen, waarbij deze suggesties hebben gedaan ter verbetering van de taal, zodat deze sneller/beter begrepen kon worden en daarmee dus intuïtiever is geworden. Door het begrip intuïtiviteit te voorzien van een meetbare score, kan zichtbaar worden gemaakt dat deze verbetering zich ook daadwerkelijk en significant heeft voorgedaan.

Hieruit volgen de volgende hypothesen:

- H1: Interpretatieve effectiviteit van de PGA-2 notatie voor element X tijdens taak Y > Interpretatieve effectiviteit PGA-1 notatie voor element X tijdens taak Y
- H2: Interpretatieve efficiëntie PGA-2 notatie voor element X tijdens taak Y < Interpretatieve efficiëntie PGA-1 notatie voor element X tijdens taak Y
- H3: Ervaren intuïtiviteit PGA-2 notatie voor element X > Ervaren intuïtiviteit PGA-1 notatie voor element X

3.2.5. Respondenten

Zoals gezegd is het van belang dat de invloed van voorkennis van zowel taal als domein zo beperkt mogelijk is. Het gewenste profiel van de respondenten ziet er als volgt uit:

- Werkzaam in de publieke sector of financiële dienstverlening
- Modelleerervaring bij voorkeur beperkt of niet aanwezig
- Professional op het gebied van (software)ontwikkeling en proces(her)ontwerp (i.e. developers, informatie- en businessanalisten, testers, *productowners*, proceseigenaren)

Deze elementen zullen worden uitgevraagd voorafgaand aan de experimentele taken. De respondenten worden vooraf niet getraind in de PGA-taal.

Toewijzing aan groep A of groep B en daarmee de volgorde van blootstelling aan de PGA-variant gebeurt op basis van geboortjaar (even vs. oneven geboortjaar). Het experiment zal volledig online worden uitgevoerd. Dit heeft als reden om deelnamedrempels te verlagen en tijd objectief te kunnen meten. Daarnaast speelt de invloed van COVID-19 op het experiment hierin een belangrijke rol. Deelname aan het onderzoek en de verwerking van de resultaten anoniem wordt uitgevoerd. De respondenten krijgen evenmin vooraf te horen tot welke groep zij behoren.

3.3. Gegevensanalyse

We hebben te maken met een aantal tweetal objectieve afhankelijke variabelen en een subjectieve afhankelijke variabele, te weten:

- de interpretatieve effectiviteit, gevormd door juiste antwoorden;
- de interpretatieve efficiëntie, gevormd verstreken tijd
- de ervaren intuïtiviteit

In het geval van interpretatieve effectiviteit hebben we te maken met numerieke, gepaarde, ordinale data. In het geval van interpretatieve efficiëntie is dit niet het geval, daar hebben we te maken met continue, gepaarde data.

De twee groepen tezamen is kleiner dan 50 respondenten. Om die reden wordt getest of de data zich normaal gedraagt met de *Shapiro-Wilk* test.

Wanneer de data zich normaal gedraagt kan worden volstaan met de t-Toets. Deze test of de gemiddelden van twee groepen van elkaar verschillen. Gedraagt de data zich niet normaal, dan zal de *Wilcoxon-Mann-Whitney* toets worden uitgevoerd om het verschil in medianen tussen de twee groepen te toetsen.

In geval van de ervaren intuïtiviteit hebben we te maken met ongepaarde data. Wanneer deze data zich normaal gedraagt zal de *One-sample t-Test* worden uitgevoerd. Gedraagt deze zich niet-normaal, dan zal de *One-sample Wilcoxon Signed Rank* test worden uitgevoerd.

3.4. Reflectie t.a.v. validiteit, betrouwbaarheid en ethische aspecten

Om interne validiteit zoveel mogelijk te waarborgen is geprobeerd de omstandigheden voor de respondenten zo vergelijkbaar mogelijk te maken door leereffecten te voorkomen met de toepassing van een *mixed design*, waarbij de respondenten binnen de experimentele taken worden geconfronteerd met *between-subjects* - en een *within-subjects* methoden om data te verzamelen.

Onder de redenen om voor een gemengd onderzoeksontwerp te kiezen vallen onder andere triangulatie, het combineren van data om vast te stellen of de bevindingen van methode A. onderling overeenkomen met de bevindingen van methode B, en het kunnen hebben van vertrouwen in de conclusies. Wanneer er slechts één methode wordt gebruikt is het niet vast te stellen of de gebruikte methode de bevindingen heeft beïnvloed. Met een gemengde methode kan een eventueel effect door de gebruikte methode wel worden vastgesteld.

Daarnaast wordt met name de voorkennis met betrekking tot het domein gemitigeerd of geëlimineerd. De PGA – taal op zichzelf is voor elke respondent onbekend, echter, kan er wel sprake zijn van bekendheid of ervaring met modelleertalen in het algemeen. Om dit effect zichtbaar te maken en te kunnen analyseren wordt eveneens gevraagd naar de ervaring met of kennis van modelleertalen.

Om ervoor te zorgen dat, ondanks de beperkte groepsgrootte, de resultaten zo generaliseerbaar mogelijk zijn binnen een bepaalde context, wordt de modelleerervaring van de respondenten uitgevraagd. Daarnaast wordt voor de praktische bedrijfscontext expliciet gekozen voor een die buiten de ervaringswereld (financiële dienstverlening of publieke sector) van de respondenten ligt. Wel wordt deze bedrijfscontext, in de vorm van een online maaltijdbesteldienst, gezocht waar eenieder wel een beeld bij kan hebben en er dus matige (voor)kennis van de bedrijfscontext bestaat. Door deze expliciete keuze kunnen resultaten beter veralgemeniseerd worden, omdat zij niet specifiek binnen de dagelijkse werkelijkheid van de respondenten valt.

Wat hierin een bedreiging voor de externe validiteit kan betekenen is het feit dat de respondenten nu niet binnen de gekozen bedrijfscontext werkzaam zijn, echter dit biedt geen garantie dat zij daarin nooit werkzaam zijn geweest. Ten tweede kan een bedreiging gevormd worden doordat de respondenten zichzelf scoren waar het gaat om modelleerervaring, waarin het risico bestaat dat de zij hun modelleerervaring te hoog of te laag inschatten.

Omdat er meerdere onderzoekers zijn die proberen dezelfde vraagstelling te beantwoorden en dit onderzoek onafhankelijk van elkaar uitvoeren, worden de datasets met elkaar gedeeld om zodoende de resultaten uit de eigen, relatief kleine dataset te kunnen toetsen aan de gecombineerde resultaten van alle onderzoekers. Hiermee is eveneens sprake van triangulatie, wat de betrouwbaarheid van het onderzoek ten goede komt.

Vanuit de literatuur is o.a. gezocht naar bestaande methoden en eenheden waarmee een subjectief begrip als “intuïtiviteit” objectief meetbaar gemaakt kon worden. De reden hiervoor is dat de gebruikte methoden ook daadwerkelijk datgene moeten meten wat wordt onderzocht. Door gebruik te maken van, vanuit wetenschappelijke basis, eerder verkregen inzichten met betrekking tot het vergelijken en evalueren van DSML's, intuïtiviteit in het algemeen en intuïtiviteit van visuele notaties in het algemeen kan worden zeker gesteld dat het onderzochte ook daadwerkelijk wordt gemeten.

De keuze voor een online experiment is gezien o.a. COVID-19 noodzakelijk. Echter, dit maakt wel dat de condities waaronder de respondenten deelnemen aan het experiment voor de onderzoeker niet controleerbaar of manipuleerbaar zijn. Dit kan gevolgen hebben voor de performance van de respondenten door bijvoorbeeld willekeurige fouten in de uitvoering van de experimentele taken en daarmee voor de betrouwbaarheid van de uitkomsten van het onderzoek.

Om betrouwbaarheid zoveel mogelijk te garanderen wordt data verzameld op basis van een duidelijk beschreven onderzoeksmethode. Dit maakt het onderzoek navolgbaar en repliceerbaar. Ten laatste worden zowel objectieve als subjectieve afhankelijke variabelen gebruikt om het begrip van de modellen te meten. Hierbij zijn de interpretatieve effectiviteit en interpretatieve efficiëntie objectief meetbaar in termen van goed en fout of in een tijdseenheid. De afhankelijke variabele “Ervaren Intuïtiviteit” betreft een persoonlijke en daarmee subjectieve waardering door de respondent.

Door zich aan de Nederlandse Gedragscode voor Wetenschappelijke Integriteit te conformeren wordt ethische en integere uitvoering van het onderzoek gewaarborgd. Hiertoe behoort het gebruik van wetenschappelijke methoden voor het zorgvuldig ontwerpen, uitvoeren en verslaan van het onderzoek. Op deze wijze worden ook transparantie, eerlijkheid en onafhankelijkheid geborgd.

Ethiek bestaat ook uit het zorgvuldig omgaan met de (persoons)gegevens van de respondenten. Hoewel er bedreigingen zijn voor de performance van de respondenten bij de keuze voor een online tool, wordt bij de keuze voor een online tool beter en objectief gewaarborgd dat deelname aan het experiment ook daadwerkelijk anoniem is. In het kader van het onderzoek worden een aantal vragen gesteld met betrekking tot voorkennis en ervaring, dit betreffen echter in geen geval bijzondere

persoonsgegevens als bedoeld in AVG. Hierover werden de respondenten vooraf geïnformeerd en is expliciet om hun toestemming gevraagd voor gebruikmaking van de resultaten van hun deelname aan het onderzoek.

Ten laatste is de onderzoeker zich bewust van de verantwoordelijkheid die zij draagt met betrekking tot eenieder die betrokken is bij het onderzoek en is de wetenschappelijke en maatschappelijke relevantie van dit onderzoek beschreven.

4. Resultaten

4.1. Participatie

Zoals in hoofdstuk 3 aangegeven heeft de wereldwijde uitbraak van COVID-19 gevolgen gehad voor de uitvoering van het onderzoek. Als gevolg hiervan is het onderzoek volledig online uitgevoerd onder professionals die grotendeels werkzaam zijn in de publieke sector of financiële dienstverlening. Hierbij zijn de geselecteerde respondenten per e-mail benaderd en hebben zij een deadline van 4 weken gekregen om aan het onderzoek mee te werken. Na twee weken hebben alle respondenten een reminder ontvangen. Het aantal volledig ingevulde vragenlijsten op de dag dat de deadline eindigde is uiteindelijk de basis voor dit onderzoek geworden. Uit de theorie is gebleken dat, in geval van een matige domeinkennis, een visuele notatie leidt tot een sneller en beter begrip van het domein. Om die reden is voor de casus gekozen voor een praktische bedrijfscontext waarbij de respondenten zich wel een voorstelling konden maken, maar die ver genoeg van hun dagelijkse praktijk af stond.

Het onderzoek is uitgevoerd door drie verschillende onderzoekers met vergelijkbare respondentengroepen. Voor het onderzoek van de schrijver van dit rapport zijn 19 potentiële respondenten benaderd. Hiervan hebben 17 respondenten gereageerd. Van deze reacties waren er twee reacties onvolledig en daarmee onbruikbaar. De resultaten van de overige 15 respondenten zijn voor dit onderzoek gebruikt. De resultaten van de verschillende onderzoekers zijn samengevoegd om zo de individuele onderzoeksresultaten met elkaar te kunnen vergelijken en de resultaten van de individuele onderzoeken te kunnen toetsen aan het geheel. Over de drie onderzoekers heen is het onderzoek uitgevoerd onder 44 respondenten.

Aan het onderzoek zelf liggen een drietal hypothesen ten grondslag. De nu volgende paragrafen gaan in op de wijze van toetsing en per hypothese op de resultaten van de betreffende toetsing.

4.2. H1: Interpretatieve effectiviteit van de PGA-2 notatie voor element X tijdens taak Y > Interpretatieve effectiviteit PGA-1 notatie voor element X tijdens taak Y

De interpretatieve effectiviteit wordt gemeten in juiste antwoorden, gegeven over een drietal verschillende taken, bestaande uit respectievelijk notatie-associatie, *comprehensionquestions* en *recallquestions*. Hoe hoger de score, hoe meer juiste antwoorden er zijn gegeven. Deze scores vervolgens geven per symbool aan welke variant hoger scoort, PGA-2 of PGA-1.

Deze scores zijn vervolgens onderworpen aan statistische tests om zodoende de in hoofdstuk 3 gestelde hypothesen te kunnen toetsen, te weten:

- H1: Interpretatieve effectiviteit van de PGA-2 notatie voor element X tijdens taak Y > Interpretatieve effectiviteit PGA-1 notatie voor element X tijdens taak Y
- H2: Interpretatieve efficiëntie PGA-2 notatie voor element X tijdens taak Y < Interpretatieve efficiëntie PGA-1 notatie voor element X tijdens taak Y
- H3: Ervaren intuïtiviteit PGA-2 notatie voor element X > Ervaren intuïtiviteit PGA-1 notatie voor element X

In de eerste plaats is getest of er sprake was van een normale of niet-normale verdeling van de scores. Afhankelijk daarvan is de *Wilcoxon-Man-Whitneytest* of de *t-Test* uitgevoerd op alle scores waar het gaat om gepaarde data om de verschillen en de significantie van de verschillen tussen de

notaties te bepalen. Waar het gaat om ongepaarde data is de *One-sample Wilcoxon Signed Rank* resp. de *One-sample t-Test* uitgevoerd. De volledige tests zijn terug te vinden in bijlage 4 voor de individuele dataset en bijlage 5 voor de gecombineerde dataset. De volledige dataset is bijgevoegd als separaat document bij dit rapport.

In de volgende paragrafen worden respectievelijk de resultaten beschreven van de testen van de Notatie-Associatie, de *Comprehension Questions*, de *Recall Questions* en de ervaren intuïtiviteit. Deze resultaten laten zien hoe de varianten van de notatie (PGA-1 of PGA-2) scoren per element.

4.2.1. Notatie-associatie

Met de resultaten van het onderzoek wordt de hypothese H1 getest:

De Interpretatieve effectiviteit van de PGA-2 notatie voor element X tijdens de notatie-associatie > Interpretatieve effectiviteit PGA-1 notatie voor element X tijdens notatie-associatie

	Notatie-Associatie	
	Individuele dataset	Gecombineerde dataset
Internal Goal	PGA-2 < PGA-1	PGA-2 > PGA-1
Value Proposition	PGA-2 = PGA-1	PGA-2 > PGA-1
Competence	PGA-2 = PGA-1	PGA-2 > PGA-1
Valuestream	PGA-2 < PGA-1	PGA-2 < PGA-1
Customer Goal	PGA-2 < PGA-1	PGA-2 < PGA-1

Tabel 4.1 resultaten Notatie-Associatie – Testen hypothese H1

Binnen de individuele set blijkt de PGA-2 variant gelijk of lager te scoren dan PGA-1. Dit betekent dat de hypothese op grond van de individuele dataset niet aanvaard kan worden. Uit de statistische tests blijkt echter niet dat de scores binnen de individuele dataset als significant kunnen worden aangemerkt. Dit betekent dat toeval niet kan worden uitgesloten.

Aanvullend worden de resultaten uit de individuele dataset getoetst aan de resultaten binnen de gecombineerde dataset. Wanneer we de resultaten uit de individuele dataset vergelijken met die uit de gecombineerde dataset van alle onderzoekers, dan blijkt het beeld vanuit de individuele dataset, met uitzondering van de symbolen voor Valuestream en Customer Goal niet bevestigd te worden.

De scores voor de elementen *Internal Goal*, *Value Proposition* en *Competence* zijn hoger voor de PGA-2 notatie dan voor de PGA-1 notatie. Hiermee lijkt de hypothese voor deze drie symbolen bevestigd te worden. Ook hier blijkt uit de statistische tests geen significantie in de resultaten voor deze drie symbolen.

Voor de symbolen voor *Valuestream* en *Customer goal* geldt dat de PGA-2 notatie ook in de gecombineerde dataset lager scoort dan de PGA-2 notatie.

Uitsluitend de scores van het symbool voor *Valuestream* in de gecombineerde dataset blijken significant te zijn, hetgeen betekent dat alleen voor dit symbool met zekerheid gesteld kan worden dat de score niet op toeval berust en daarmee gegeneraliseerd kan worden.

4.2.2. Comprehension-taak

De taak die de begripsvorming omvat heeft als enige taak zes symbolen uitgevraagd. Het extra symbool, *importance*, is zowel in PGA-1 als in PGA-2 gecombineerd met het symbool voor *valuestream*.

Ook binnen deze taak wordt de hypothese H1 getest:

De Interpretatieve effectiviteit van de PGA-2 notatie voor element X tijdens de comprehension-taak > Interpretatieve effectiviteit PGA-1 notatie voor element X tijdens comprehension-taak

	Comprehension	
	Individuele dataset	Gecombineerde dataset
Customer Goal	PGA-2 = PGA-1	PGA-2 < PGA-1
Value Proposition	PGA-2 < PGA-1	PGA-2 < PGA-1
Valuestream	PGA-2 > PGA-1	PGA-2 < PGA-1
Competence	PGA-2 = PGA-1	PGA-2 < PGA-1
Internal Goal	PGA-2 < PGA-1	PGA-2 < PGA-1
Importance *	PGA-2 < PGA-1	PGA-2 = PGA-1

* dit symbool is alleen uitgevraagd binnen de begripstaak

Tabel 4.2: resultaten *Comprehension*-taak – Testen hypothese H1

Binnen de individuele set blijkt de PGA-2 wisselend te scoren t.o.v. PGA-1. Er ontstaat geen eenduidig beeld. Dit betekent dat de hypothese op grond van de individuele dataset niet zonder meer aanvaard of verworpen kan worden. Uit de statistische tests blijkt daarnaast evenmin dat de scores binnen de individuele dataset als significant kunnen worden aangemerkt. Dit betekent dat toeval niet kan worden uitgesloten.

Ook hier worden de individuele resultaten getoetst aan de gecombineerde resultaten van alle onderzoekers. Wanneer we naar deze gecombineerde dataset kijken komt in vergelijking met de individuele set een meer coherent beeld naar voren. Hier blijkt PGA-2, met uitzondering van *importance*, lager te scoren dan PGA-1. Hierbij moet echter wel worden aangetekend dat de waargenomen verschillen in de data marginaal blijken te zijn (zie hiervoor ook bijlagen 4 en 5). Ook hier geldt dan ook dat er geen sprake is van significantie in de uitkomsten en kan toeval niet worden uitgesloten. Op basis van de resultaten kan de hypothese dan ook niet worden aanvaard of verworpen.

4.2.3. Recall-taak

De derde taak betrof het zich kunnen herinneren van een element en haar betekenis. Net als binnen de notatie-associatie en de taak omtrent begripsvorming wordt ook hier o.a. hypothese H1 getest:

De Interpretatieve effectiviteit van de PGA-2 notatie voor element X tijdens de recall-taak > Interpretatieve effectiviteit PGA-1 notatie voor element X tijdens recall-taak

	Recall	
	Individuele dataset	Gecombineerde dataset
Internal Goal	PGA-2 > PGA-1	PGA-2 < PGA-1
Value Proposition	PGA-2 < PGA-1	PGA-2 < PGA-1
Competence	PGA-2 > PGA-1	PGA-2 > PGA-1
Valuestream	PGA-2 < PGA-1	PGA-2 < PGA-1
Customer Goal	PGA-2 < PGA-1	PGA-2 < PGA-1

Tabel 4.3: resultaten Recall-taak – Testen hypothese H1

Binnen de *recall*-taak zijn de resultaten in de individuele dataset wisselend. Ook hier geldt dat er geen significante verschillen aangetoond kunnen worden. Dit betekent dat het verwachte effect zich niet of onvoldoende heeft voorgedaan en dat de hypothese H1 aanvaard nog verworpen kan worden waar het gaat om de individuele dataset.

Wanneer we deze toetsen aan de gecombineerde dataset blijken de resultaten nagenoeg gelijk aan elkaar, met uitzondering van het element *Internal Goal*. Hier is sprake van aan elkaar tegengestelde resultaten. Hoewel ook hier de geconstateerde verschillen niet significant zijn en er dus geen sprake kan zijn van het aanvaarden of verwerpen van hypothese H1, lijken de resultaten elkaar voor vier van de vijf elementen wel te bevestigen. Dit zou echter nader onderzocht moeten worden.

4.3. H2: Interpretatieve efficiëntie PGA-2 notatie voor element X tijdens taak Y < Interpretatieve efficiëntie PGA-1 notatie voor element X tijdens taak Y

Naast de scores op (gedeeltelijke) juistheid is ook de tijd gemeten die de respondenten nodig hadden voor het oplossen van de taak om de efficiëntie waarmee zij de taken afrondden te kunnen bepalen. Hiermee wordt hypothese H2 getest:

De interpretatieve efficiëntie PGA-2 notatie voor element X tijdens taak Y < Interpretatieve efficiëntie PGA-1 notatie voor element X tijdens taak Y

Hiervoor zijn eveneens statistische testen uitgevoerd om de normaliteit van de resultaten en de significantie van geconstateerde verschillen te kunnen bepalen. De statistische testen die hiervoor zijn gebruikt zijn gelijk aan de testen gebruikt voor het testen van H1.

Hierbij moet worden opgemerkt dat wanneer het resultaat van PGA-2 kleiner is dan dat van PGA-1, dit betekent dat er minder tijd nodig was voor het oplossen van de betreffende vraag en er dus efficiënter tot een antwoord gekomen kon worden.

De gemeten tijden zijn alleen relevant in geval van de eerste drie taken, notatie-associatie, *comprehension*-taak en *recall*-taak. Aan deze vragen is immers een goed of fout score verbonden. Deze geldt niet voor de ervaren intuïtiviteit. De resultaten van de testen en de daarbij geconstateerde verschillen zijn terug te vinden in onderstaande tabellen 4.4 tot en met 4.6.

4.3.1. Notatie-associatie

Zoals aangegeven wordt met dit deel van het onderzoek hypothese H2 getest:

De interpretatieve efficiëntie PGA-2 notatie voor element X tijdens de notatie-associatie < Interpretatieve efficiëntie PGA-1 notatie voor element X tijdens de notatie-associatie

	Notatie-Associatie tijd	
	Individuele dataset	Gecombineerde dataset
Internal Goal	PGA-2 < PGA-1	PGA-2 < PGA-1
Value Proposition	PGA-2 > PGA-1	PGA-2 > PGA-1
Competence	PGA-2 > PGA-1	PGA-2 > PGA-1
Valuestream	PGA-2 < PGA-1	PGA-2 < PGA-1
Customer Goal	PGA-2 > PGA-1	PGA-2 > PGA-1

Tabel 4.4: Notatie Associatie— Testen hypothese H2

De individuele dataset laat een wisselend beeld zien. Niet in alle gevallen is interpretatieve efficiëntie groter in het geval van PGA-2. Daarnaast blijken de geconstateerde verschillen niet significant te zijn. H2 kan hiermee dan ook niet verworpen of aanvaard worden.

In de toetsing van de individuele dataset aan de gecombineerde dataset valt op dat de resultaten aan elkaar gelijk zijn. Hiermee lijken de resultaten uit de gecombineerde dataset die van de individuele dataset te bevestigen. De elementen *Internal Goal* en *Valuestream* scoren lager bij PGA-2 dan bij PGA-1. Dit betekent dat voor een associatie minder tijd nodig is bij deze symbolen. In alle overige situaties geldt dat er meer tijd nodig is voor een associatie in PGA-2 dan in PGA-1.

Ook hier zijn echter de geconstateerde verschillen tussen PGA-2 en PGA-1 niet significant, waarmee toeval dan ook niet kan worden uitgesloten.

4.3.2. Comprehension-taak

Binnen de *Comprehension*-taak wordt eveneens naar de interpretatieve efficiëntie gekeken en hypothese H2 getest:

De interpretatieve efficiëntie PGA-2 notatie voor element X tijdens de Comprehension-taak < Interpretatieve efficiëntie PGA-1 notatie voor element X tijdens de Comprehension-taak

	Comprehension tijd	
	Individuele dataset	Gecombineerde dataset
Customer Goal	PGA-2 > PGA-1	PGA-2 > PGA-1
Value Proposition	PGA-2 < PGA-1	PGA-2 < PGA-1
Valuestream	PGA-2 > PGA-1	PGA-2 > PGA-1
Competence	PGA-2 < PGA-1	PGA-2 < PGA-1
Internal Goal	PGA-2 > PGA-1	PGA-2 < PGA-1
Importance *	PGA-2 < PGA-1	PGA-2 < PGA-1

* dit symbool is alleen uitgevraagd binnen de begripstaak

Tabel 4.5 : Comprehension-taak – Testen hypothese H2

Ook hier geldt dat de individuele dataset een wisselend beeld laat zien. Dit wisselende beeld laat zich ook zien in de gecombineerde dataset. Met uitzondering van de resultaten voor het element *internal Goal* zijn in zowel de individuele- als de gecombineerde dataset dezelfde verschillen te constateren. Binnen beide datasets zijn uitsluitend de geconstateerde verschillen van het symbool *Importance* significant gebleken.

Dit betekent dat zowel in geval van de individuele dataset als in de gecombineerde dataset hypothese H2 in zijn geheel aanvaard nog verworpen kan worden.

Alleen voor *Importance* kan worden gesteld dat de betere score voor interpretatieve efficiëntie binnen PGA-2 niet op toeval berust. Het symbool voor wordt in verder gebruik gecombineerd met het symbool voor *Valuestream* in zowel PGA-1 als PGA-2. *Importance* wordt in verdere taken niet specifiek uitgevraagd en maakt daarmee vergelijking niet mogelijk.

4.3.1. Recall-taak

Ook voor de *recall*-taak geldt dat met de resultaten de hypothese H2 wordt getoetst:

De interpretatieve efficiëntie PGA-2 notatie voor element X tijdens de Recall-taak < Interpretatieve efficiëntie PGA-1 notatie voor element X tijdens de recall-taak

Ook hier laat de individuele dataset een sterk wisselend beeld zien en zijn de geconstateerde verschillen niet significant. Vanuit de individuele dataset kan ook in het geval van deze taak de hypothese H2 aanvaard noch verworpen worden.

Wat opvalt in de gecombineerde dataset, is dat de resultaten grotendeels tegenovergesteld zijn aan die van de individuele dataset. Dit blijft evenwel eveneens een wisselend beeld tonen.

De geconstateerde verschillen tussen PGA-2 en PGA-1 zijn significant voor de elementen *Internal Goal* en *Competence*. Hiermee kan de hypothese H2 voor deze twee specifieke symbolen weliswaar worden aanvaard, echter, dit geldt niet voor alle symbolen.

	Recall tijd	
	Individuele dataset	Gecombineerde dataset
Internal Goal	PGA-2 > PGA-1	PGA-2 < PGA-1
Value Proposition	PGA-2 < PGA-1	PGA-2 < PGA-1
Competence	PGA-2 > PGA-1	PGA-2 < PGA-1
Valuestream	PGA-2 > PGA-1	PGA-2 > PGA-1
Customer Goal	PGA-2 < PGA-1	PGA-2 > PGA-1

Tabel 4.6 : Recall-taak – Testen hypothese H2

4.4. H3: Ervaren intuïtiviteit PGA-1 < Ervaren intuïtiviteit PGA-2

In de vierde en laatste taak is de respondenten gevraagd aan te geven welke variant zij als meer intuïtief ervaren. Hierbij kregen de respondenten beide varianten te zien, echter is er rekening gehouden met welke variant zij in de drie eerdere taken hebben gewerkt. Deze variant stond steeds als eerste keuze vermeld.

Zowel in de individuele dataset als in de gecombineerde dataset is grotendeels een voorkeur voor PGA-2 aangegeven. Binnen de individuele dataset geldt dat voor het element *Competence* de voorkeur is aangegeven voor PGA-1. In de gecombineerde dataset is een gelijke score geconstateerd voor PGA-1 en PGA-2 wanneer het gaat om het symbool voor *Valuestream*.

De resultaten van de individuele dataset zijn niet significant. Voor de gecombineerde dataset geldt dat de resultaten voor de elementen *Customer Goal* en *Value Proposition* significant zijn. De overige resultaten zijn niet significant.

	Representatievoorkeur	
	Individuele dataset voorkeur	Gecombineerde dataset voorkeur
Customer Goal	PGA-2	PGA-2
Value Proposition	PGA-2	PGA-2
Competence	PGA-1	PGA-2
Internal Goal	PGA-2	PGA-2
Valuestream	PGA-2	PGA-2 = PGA-1

Tabel 4.7: Representatievoorkeur – Testen H3

Een aantal citaten met betrekking tot de notaties geven hierbij meer inzicht.

Met betrekking tot het element *Customer Goal* geeft een respondent aan:

“Slight preference for A as it shows there is a customer (without brief case) and a provider of Services (with brief case). However, both have a too strong association for me with a contractual agreement.”

In geval van het element *Value Proposition* worden de volgende opmerkingen (of een variant daarop) terug gezien:

“B reminds me of Justice”

En

“Even though value isn't necessary free of cost (which the present implies), it is at least related to delivering something. Undelivered value isn't value.”

Voor element *Competence* was de algemene consensus dat geen van beide varianten als erg intuïtief werden ervaren. In de individuele dataset scoort dit element tegengesteld aan het symbool in PGA-2.

Een respondent gaf hierover de volgende opmerking :

“A light bulb typically is related to an idea. Competence is something you are able to do.

Competence is also not a competition, regardless of the similarity of the words. Both symbols are equally unsuitable. I would much rather see a diploma-like symbol or a person with a graduation hat.”

4.5. Geaggregeerde scores

Om de elementen met elkaar te kunnen vergelijken over de taken heen zijn de scores van alle taken omgezet naar een percentage. Vervolgens zijn hier eveneens de eerdergenoemde statistische tests

voor uitgevoerd. De volledige tests zijn bijgevoegd in bijlage 6. De oorspronkelijke data is als separaat document bijgevoegd bij dit rapport.

De waargenomen verschillen zijn in de individuele dataset en de gecombineerde dataset gelijk aan elkaar. Opvallend is dat voor vrijwel alle symbolen geldt dat de PGA-1 notatie beter scoort. Slechts het symbool voor *Customer Goal* scoort beter in de PGA-2 notatie. Ook hier geldt echter dat de waargenomen verschillen niet significant zijn.

	Geaggregeerde scores	
	Individuele dataset	Gecombineerde dataset
Internal Goal	PGA-2 < PGA-1	PGA-2 < PGA-1
Value Proposition	PGA-2 <PGA-1	PGA-2 < PGA-1
Competence	PGA-2 <PGA-1	PGA-2 < PGA-1
Value Stream	PGA-2 < PGA-1	PGA-2 < PGA-1
Customer Goal	PGA-2 > PGA-1	PGA-2 >PGA-1

Tabel 4.8: resultaten geaggregeerde scores notatie

5. Discussie, conclusies en aanbevelingen

In dit hoofdstuk wordt inhoudelijk ingegaan op het uitgevoerde onderzoek en de in hoofdstuk 4 gepresenteerde onderzoeksresultaten. Achtereenvolgens zullen deze worden geïnterpreteerd en verklaringen worden gezocht en, indien mogelijk, conclusies aan worden verbonden. Ten laatste worden aanbevelingen naar aanleiding van het onderzoek gedaan.

5.1. Discussie – reflectie

5.1.1. Validiteit

Met dit onderzoek heeft de onderzoeker de mate van intuïtiviteit van de PGA-notaties willen meten, om zodoende te kunnen bepalen of er zich een verbetering van de intuïtiviteit van de verbeterde PGA-notatie (PGA-2) heeft voorgedaan ten opzichte van de oorspronkelijke notatie (PGA-1). Hierbij is als maatstaven van intuïtiviteit aangehouden: Interpretatieve effectiviteit en interpretatieve efficiëntie enerzijds en de ervaren intuïtiviteit anderzijds. Deze twee elementen komen ook in andere onderzoeken keer op keer naar voren als maatstaf voor begripsvorming. Zoals gebleken uit de literatuur is begripsvorming een sleutelement binnen de te meten intuïtiviteit. Door de mate van bekendheid met de notatie te manipuleren (summiere informatie voorafgaand aan het onderzoek) en het betreffende domein buiten de dagelijkse praktijk van de respondenten te stellen wordt voldaan aan beperkte voorkennis.

Triangulatie is toegepast in de vorm van onderzoeks-triangulatie om de validiteit te kunnen waarborgen. Hierbij is het onderzoek door meerdere onderzoekers onafhankelijk van elkaar uitgevoerd, elk met hun eigen groep respondenten om de hoofdvraag te beantwoorden. Na uitvoering van het onderzoek zijn de resultaten van de verschillende onderzoekers samengevoegd.

Het onderzoek is opgezet als experiment. Hoewel bij de selectie van de respondenten rekening is gehouden met de functieprofielen van de respondenten zijn deze uiteindelijk niet uitgevraagd. Dit is gebeurd om te voorkomen dat antwoorden zouden zijn te herleiden naar de individuele respondenten en daarmee de anonimiteit in gevaar gebracht kon worden.

Ook opleidingsrichting is in eerste instantie genoemd als middel om voorkennis m.b.t. het domein te mitigeren. Eveneens uit privacyoverwegingen is dit uiteindelijk niet uitgevraagd. Wel is vastgehouden aan de eis dat de groep respondenten zou bestaan uit professionals op het gebied van (software)ontwikkeling en/of proces(her)ontwerp. Hiermee zijn de onderzoeksgroepen over de onderzoekers heen nog steeds vergelijkbaar. Wel is mogelijk de wijze van selectie van respondenten verschillend over de onderzoekers heen. Hiermee kan in die zin de interne validiteit worden bedreigd doordat er nadruk kan liggen op het type professional dat is benaderd. Dit kan in de individuele onderzoeksresultaten zorgen voor verschillen in resultaten. Deze worden echter door het toepassen van onderzoeks-triangulatie geëgaliseerd.

5.1.2. Betrouwbaarheid

Zoals eerder aangegeven is geprobeerd de omstandigheden voor de respondenten zo vergelijkbaar mogelijk te maken. Binnen de experimentele taken zijn de respondenten in beide groepen blootgesteld aan dezelfde opdrachten en een mix van PGA-1 en PGA-2. Hiermee is de onderzoeksmethode voor beide groepen en voor elke respondent steeds in dezelfde setting geweest en de methode structureel volgens exact dezelfde procedures uitgevoerd. De tijd tussen de taken is gemeten door de LimeSurvey applicatie waarmee de respondenten hebben deelgenomen aan het onderzoek. Dit betekent dat er geen menselijke meetfouten kunnen zijn gemaakt. Ook houdt dit in

dat de meetmethode tussen zowel de respondenten als de onderzoekers niet kan afwijken. Die is in alle gevallen hetzelfde geweest.

Een kanttekening moet hierbij overigens wel worden gemaakt. Omdat er volledig onlineonderzoek is uitgevoerd, was het voor de onderzoeker niet mogelijk om invloed uit te oefenen op de omgeving waarin de respondenten aan het onderzoek hebben deelgenomen. kan niet worden uitgesloten dat respondenten werden afgeleid door hun omgeving op het moment van deelname. Om die reden moet dan ook met enige reserve gekeken worden naar de besteedde tijd per vraag wanneer het gaat om individuele datasets. In de gecombineerde datasets zullen eventuele afwijkingen worden geëgaliseerd door de groepsgrootte.

Ondanks deze kanttekening kan op grond van de methodiek en meetinstrumenten wel worden gesteld dat het onderzoek betrouwbaar is uitgevoerd. Daarnaast is, op grond van de beschreven methodiek, het onderzoek navolgbaar en herhaalbaar. Ook dit komt de betrouwbaarheid ten goede.

5.1.3. Interpretatieve effectiviteit

De interpretatieve effectiviteit wordt gemeten door te kijken naar de juiste antwoorden op de gestelde vragen.

Wanneer het gaat om het begrip van het element Valuestream binnen een casus, zo blijkt uit de data, scoort PGA-2 minder dan PGA-1. Dit patroon is zichtbaar binnen de gehele gecombineerde dataset wanneer het gaat om begrip binnen een casus en zet zicht voort bij de herinneringsopdracht en is terug te zien in de geaggregeerde scores voor de elementen over de opdrachten heen gemeten. Hoewel de verschillen tussen beide varianten overwegend niet significant blijken te zijn, wordt hiermee wel meer een patroon zichtbaar binnen een grotere groep respondenten. Binnen de individuele dataset zijn de resultaten meer uiteenlopend en patronen moeilijk te detecteren...

Opvallend is dat binnen de notatie-associatie zichtbaar wordt dat de elementen *Value Proposition* en *Competence* een hogere score geven binnen PGA-2. Binnen de notatie-associatieopdracht waren de respondenten nog niet bekend met PGA-taal. In geval van matige domeinkennis zou een visuele notatie het meeste effect hebben op het begrip van de respondenten over het getoonde domein. Voor beide begrippen geldt dat binnen de Notatie-Associatie – opdracht PGA-2 beter scoort, terwijl binnen de casestudy met begripsvragen PGA-1 scoort voor beiden.

Voor het element Competence geldt dat, binnen de herinneringsopdracht in de gecombineerde dataset, eveneens blijkt dat PGA-2 beter scoort. Binnen de begrips-opdracht scoort dit element beter in PGA-1.

Op het eerste gezicht lijkt dit elkaar tegen te spreken. Het zou echter ook zo kunnen zijn dat de individuele symbolen wel meer intuïtief zijn voor deze twee elementen dan in de oorspronkelijke notatie, maar dat met de combinatie van de notaties binnen een casus een complexiteit wordt toegevoegd die het eerdere positieve effect tenietdoet.

Dit lijkt in het geval van het element *Competence* binnen de complete dataset te worden ondersteund. Eenzelfde patroon, een tegengestelde uitkomst binnen taak twee t.o.v. de eerste en derde taak, is ook te zien in de overige elementen.

De invloed van de complexiteit van een taak behoort niet tot de scope van dit onderzoek. Mocht de complexiteit een invloed van betekenis hebben, dan heeft dit gevolgen voor het gebruik van de PGA-taal in de praktijk. Immers, het doel is om de processen en doelstellingen van een onderneming

visueel met elkaar te verbinden en zodoende een beter begrip in de vertaling van strategie naar operatie te realiseren.

Binnen de geaggregeerde scores blijkt juist het element Customer Goal beter te scoren in PGA-2. In de individuele scores is dit niet terug te zien. Hierbij moet worden opgemerkt dat binnen de geaggregeerde scores niet alleen de juiste antwoorden zijn opgenomen, maar ook de ervaren intuïtiviteit.

Ten laatste is het niet onmogelijk dat de geconstateerde verschillen tussen de notaties simpelweg te klein zijn om significant te kunnen zijn.

5.1.4. Interpretatieve efficiëntie

De interpretatieve efficiëntie wordt gemeten in de snelheid waarmee de antwoorden op de vragen worden gegeven. Zoals in de discussie aangegeven moet hierbij wel een zekere reserve worden aangehouden waar het gaat om de individuele dataset. Wanneer er sprake is van afleiding in de omgeving van de respondent zal dit zwaarder wegen dan in de gecombineerde dataset. De tijd besteed aan de representatievoorkeur is niet relevant voor het onderzoek. Ook de gecumuleerde tijd voor de geaggregeerde scores is hierin niet relevant.

Binnen de geconstateerde verschillen zijn alleen bij de *recall questions* in de gecombineerde dataset significante verschillen te benoemen en wel voor de elementen Valuestream en Customer Goal. Voor deze elementen is binnen PGA-2 meer tijd nodig om tot een antwoord te komen dan bij PGA-1.

Wat opvalt binnen het geheel is dat met name bij de derde taak er sprake is van resultaten die tegengesteld zijn aan elkaar in de vergelijking tussen de individuele en de gecombineerde dataset. Mogelijk is het herinneren van een symbool minder gerelateerd aan intuïtiviteit dan in eerste instantie gedacht. Hiernaar zou echter verder onderzoek moeten worden gedaan.

5.1.5. Ervaren intuïtiviteit

Zowel in de individuele – als in de gecombineerde dataset blijkt de voorkeur van de respondenten overwegend voor PGA-2 te worden aangegeven. Zoals uit het onderzoek van (Bobkowska, 2005) blijkt is de persoonlijke reflectie over de ervaren intuïtiviteit een belangrijk deel van het onderzoek.

Het belang van de persoonlijke reflectie blijkt uit de gedane uitspraken waarmee de voorkeur kon worden gemotiveerd. Deze geven veel extra informatie.

Hieruit blijkt ook heel duidelijk een manco binnen de gekozen manier van reflecteren. De respondenten hadden de keuze uit twee symbolen. Daarnaast was het invullen van een motivatie facultatief. Bij degenen die de motivatie wel hebben gegeven was bij een aantal respondenten de reactie “*Both icons are not very intuitive*” of een variant daarop. Dit kan betekenen dat er gekozen is voor “de minst slechte” variant. Wanneer inderdaad is gekozen voor de minst slechte variant, dan geeft dat geen garantie op een intuïtieve notatie, waarmee makkelijk begrip door organisatielagen onder druk komt te staan.

Wat we hierin niet mogen vergeten is dat dit onderzoek is uitgevoerd onder professionals. De eerdere onderzoeken zijn uitgevoerd onder studenten. Mogelijk reageren studenten bij gebrek aan ervaring met diverse andere modelleertalen anders op de PGA-notaties dan professionals die wel al gedurende kortere of langere tijd daaraan zijn blootgesteld. Mocht dit inderdaad een verschil van betekenis zijn, dan betekent dat een dergelijk onderzoek onder studenten niet zonder meer gebruikt kan worden om resultaten te generaliseren.

5.2. Conclusies

Uit het literatuur onderzoek is een basis voor onderzoek naar intuïtiviteit van modelleertalen gebleken. Deze basis is gevonden in onderzoek naar begripsvorming van notaties en het evalueren van notaties. De brug van begripsvorming naar intuïtiviteit is gevormd door de onderzoeken van Jošt et al. (2016) waarin een definitie van intuïtiviteit in relatie tot modelleertalen wordt gegeven en het onderzoek van Bork et al. (2019) en (Roelens et al., 2019) waarin expliciet een methode wordt beschreven voor het onderzoeken van de intuïtiviteit van modelleertalen. Dit laatste onderzoek is dan ook het vertrekpunt geweest voor het ontwerp van het onderzoek zoals beschreven in deze paper.

Verder is vanuit de literatuur gebleken dat een experiment met een gemengd ontwerp van *between-subjects* en *within-subject* de geëigende manier is om begripsvorming en intuïtiviteit van visuele notaties te onderzoeken.

Ook is gebleken dat door het meten van de juiste antwoorden (interpretatieve effectiviteit) en de benodigde tijd voor het oplossen van een taak (de interpretatieve efficiëntie) objectief gemeten kan worden. Daarbij is met het onderzoek van Bobkowska (2005) aangetoond dat naast de objectieve maatstaven ook de subjectieve maatstaf in de vorm van ervaren intuïtiviteit nodig is om een goed totaalbeeld te krijgen van de intuïtiviteit van een modelleertaal.

Uit de resultaten van het empirisch onderzoek is gebleken dat, hoewel er verschillen zijn waargenomen tussen de beide notaties, deze verschillen overwegend niet significant zijn. Binnen de individuele dataset is een significant verschil waargenomen, binnen de gecombineerde dataset zijn een zestal significante verschillen geconstateerd. Voor deze gevallen kan worden gesteld dat deze uitkomsten niet op toeval berusten. Omdat er in deze vergelijkende evaluatie twee verschillende varianten van dezelfde taal met elkaar zijn vergeleken, is het niet ondenkbaar dat de verschillen tussen de oorspronkelijke en de nieuwe notatie simpelweg te klein zijn om te kunnen spreken van een aantoonbare verandering.

Niet-significant houdt overigens niet in “Niet bestaand”. Dat toeval niet uitgesloten kan worden betekent niet dat er in het geheel geen conclusies getrokken kunnen worden. Uit het literatuuronderzoek is gebleken dat het evalueren van een visuele modelleertaal een niet te onderschatten onderzoek is en dat daarom de evaluatie van de respondenten een noodzakelijke, bijkomende manier is om een evaluatie uit te kunnen voeren. Dit lijkt door dit onderzoek ten dele ondersteund te worden.

5.2.1. Conclusies Hypothesen

Definitieve uitspraken met betrekking tot de hypothesen kunnen niet zonder meer gedaan worden. Wanneer we uitgaan van de onderzoeksresultaten, zonder te kijken naar de significantie van de verschillen, dan komen daaruit de volgende conclusies:

- H1: Interpretatieve effectiviteit van de PGA-2 notatie voor element X tijdens taak Y > Interpretatieve effectiviteit PGA-1 notatie voor element X tijdens taak Y

H1 is niet aangetoond kunnen worden. Uit de resultaten blijkt overwegend het tegengestelde. In de geaggregeerde resultaten is zelfs in 80% van de gevallen de score voor PGA-1 hoger. Hiermee zou de hypothese verworpen moeten worden, ware het niet dat, m.u.v. één symbool binnen de gecombineerde dataset, de verschillen niet significant zijn. Toeval kan daarmee niet worden uitgesloten

- H2: Interpretatieve efficiëntie PGA-2 notatie voor element X tijdens taak Y < Interpretatieve efficiëntie PGA-1 notatie voor element X tijdens taak Y

H2 is niet aangetoond kunnen worden in de individuele dataset. In negen van de zestien resultaten bleek dit andersom. In de gecombineerde dataset daarentegen is H2 aangetoond in 10 van de 16 resultaten. Ook hier geldt: de verschillen zijn overwegend niet significant, op een drietal uitzonderingen na. De resultaten zijn echter ook veel minder uniform dan de resultaten m.b.t. de interpretatieve effectiviteit. Zoals aangegeven in de discussie moet er met enige reserve naar deze onderzoeksresultaten worden gekeken.

- H3: Ervaren intuïtiviteit PGA-2 notatie voor element X > Ervaren intuïtiviteit PGA-1 notatie voor element X

Zowel binnen de individuele als binnen de gecombineerde dataset is overwegend een voorkeur gebleken voor de PGA-2 variant. Voor de elementen Customer Goal en Value Proposition is het verschil in voorkeur voor PGA-2 ook significant gebleken. Echter, hier kan niet worden uitgesloten, op basis van de motivaties van de respondenten dat hier is gekozen voor “de minst slechte variant”.

De conclusie die getrokken moet worden is dat dit onderzoek statistisch gezien niet aan heeft kunnen tonen dat de verbetering in PGA-2 zich daadwerkelijk heeft voorgedaan. Wat dat betekent voor de praktijk en voor verder onderzoek wordt beschreven in de volgende paragrafen.

5.3. Aanbevelingen voor de praktijk

Zoals aangegeven heeft de verbetering in PGA-2 niet aangetoond kunnen worden binnen dit onderzoek. Hierbij moet in ogenschouw genomen worden dat het verschil tussen dit onderzoek en het onderzoek waar dit op voortbouwt is gelegen in het type respondenten. Dit onderzoek is het eerste onderzoek wat is uitgevoerd onder professionals. Voor de praktijk betekent dit dat nu nog niet het moment is aangebroken om de PGA-taal te introduceren voor gebruik door professionals in hun bedrijfscontext. Hiertoe is eerst verder onderzoek noodzakelijk.

Wanneer we teruggrijpen op het oorspronkelijke onderzoek van Bork et al. (2019), waarmee de experimentele methode initieel is ontwikkeld, is het zinvol om dit onderzoek opnieuw uit te voeren gecombineerd met het onderhavige onderzoek, waarbij dit keer als respondent wordt gekozen voor professionals in plaats van studenten.

De combinatie van beide onderzoeken bestaat uit het herhalen van beide onderzoeken, volgorde-lijk in de tijd, bij voorkeur met verschillende groepen respondenten, om zodoende een leereffect te voorkomen.

Hiermee wordt de experimentele evaluatie direct toegepast in de praktijk en maakt de evaluatie en de voorstellen tot verbetering van de PGA-taal minder theoretisch. Vervolgens zal ook dan moeten worden getest of de verbetering zich daadwerkelijk voordoet.

Door te kiezen voor een populatie die praktijkervaring kent, wordt de ervaring van de respondenten belangrijk en kan van grote toegevoegde waarde zijn, immers, professionals weten wat in praktijk wel en niet werkt. Dat zou zich moeten vertalen in meer intuïtieve notatie-voorstellen en uiteindelijk een significante verbetering van de notaties.

Hoewel de praktijkervaring van professionals als respondenten een grote bijdrage kan leveren aan de praktische ontwikkeling van een meer intuïtieve notatie, moet niet worden vergeten dat dezelfde praktijkervaring ook kan leiden tot “notatieblindheid”. Hiermee doelt de onderzoeker op het feit dat

bekendheid met andere notaties kan leiden tot een wat minder onbevangen perceptie van de notaties.

5.4. Aanbevelingen voor verder onderzoek

Er zijn een aantal mogelijke oorzaken genoemd voor de uitkomsten van dit onderzoek en de implicaties van deze oorzaken. Deze punten zijn van belang om in vervolgonderzoek te toetsen.

Zo is genoemd dat het combineren van symbolen mogelijk een complexiteit binnen de taken of visuele weergave brengt die de mate van intuïtiviteit beïnvloedt. Wanneer uit vervolgonderzoek blijkt dat dit inderdaad het geval is en in welke mate dit het geval is, kan de taal daarop worden aangepast.

Een dergelijk onderzoek zou op verschillende manieren vorm gegeven kunnen worden. Er zou gedacht kunnen worden aan een opzet waarbij er binnen het onderzoek een verschil is in complexiteit van de opdrachten. Waarbij de onderzoeksgroep A wordt geconfronteerd met taken die betrekking hebben op de combinatie van symbolen en onderzoeksgroep B uitsluitend geconfronteerd wordt met taken die betrekking hebben op de individuele symbolen.

Een andere vorm zou kunnen zijn door de opdrachten op te bouwen in complexiteit, waarbij het vertrekpunt ligt op het individuele symbool en opklimt via de niveaus binnen de PGA-taal van activiteit, via proces en valuestream naar het uiteindelijke doel.

Ook is genoemd dat een herinneringstaak mogelijk niet de meest geschikte manier is om intuïtiviteit te meten gezien de resultaten van het onderzoek. Verder (literatuur)onderzoek zou dit moeten uitwijzen. Wanneer het gaat om herinnering en begripsvorming komt men al snel uit op psychologisch onderzoek en onderzoek naar de hersenen. Er wordt bijvoorbeeld veel onderzoek gedaan naar dementie wat veel inzichten kan geven in de relatie tussen begripsvorming en herinnering. Ook binnen de discipline van taalwetenschappen kunnen nieuwe inzichten worden verworven. Denk hierbij bijvoorbeeld aan de ontwikkeling van emoticons vanuit deze invalshoek.

Ten derde is genoemd dat binnen de gekozen methode bij de ervaren intuïtiviteit mogelijk een keuze is gemaakt voor “de minst slechte” optie. Om dit nader te onderzoeken zijn diverse methoden denkbaar. Zo kan de reflectie bijvoorbeeld in vervolgonderzoek plaatsvinden d.m.v. interviews. Ook zou gekozen kunnen worden voor een motivatie die niet facultatief is.

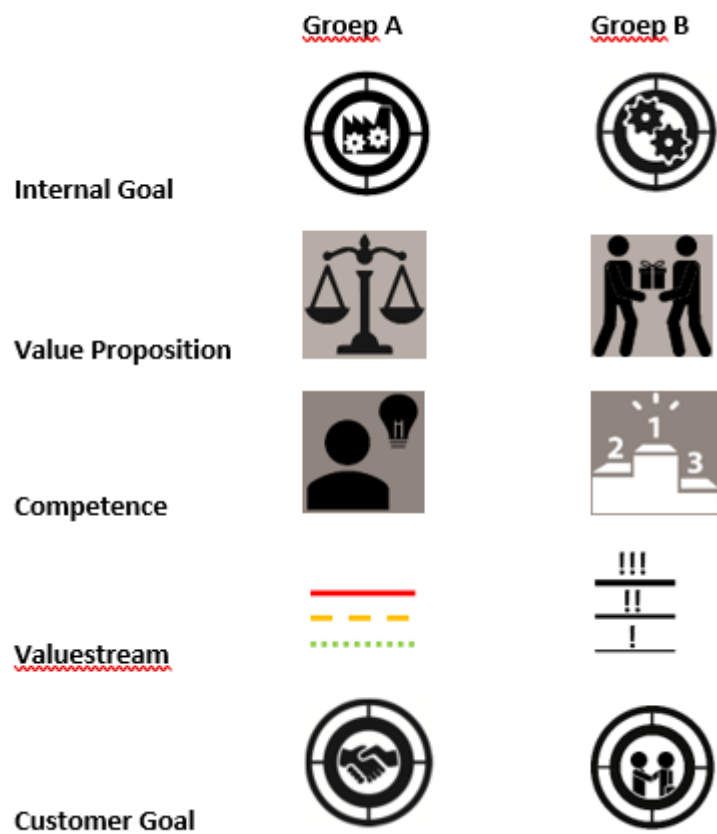
Dit onderzoek betreft een onderzoeksmethode voor het vergelijkend kunnen evalueren van de intuïtiviteit van modelleertalen, die voortkomt uit een nieuw ontwikkelde methode die specifiek bedoeld is om intuïtiviteit van een notatie te onderzoeken. Het is altijd zinvol om in een dergelijk vroeg stadium ook de methode als geheel te toetsen.

Een dergelijk onderzoek zou bijvoorbeeld kunnen bestaan door de methodiek van dit onderzoek te testen in een vergelijkende evaluatie van twee van elkaar verschillende modelleertalen. De verschillen tussen twee talen zullen naar verwachting groter zijn dan de verschillen tussen twee versies van dezelfde taal en kunnen daarmee tonen of de methodiek ook daadwerkelijk geschikt is om de intuïtiviteit van een taal te bepalen.

Referenties

- Bera, P., Burton-Jones, A., & Wand, Y. (2014). How Semantics and Pragmatics Interact in Understanding Conceptual Models. *Information Systems Research*, 25(2), 401-419. doi:10.1287/isre.2014.0515
- Bobkowska, A. (2005). *A methodology of visual modeling language evaluation*. Paper presented at the SOFSEM 2005: Theory and Practice of Computer Science, Liptovský Ján, Slovakia.
- Bork, D., Schrüffer, C., & Karagiannis, D. (2019). Intuitive Understanding of Domain-Specific Modeling Languages: Proposition and Application of an Evaluation Technique. In (pp. 311-319).
- Cachero, C., Meliá, S., & Hermida, J. M. (2019). Impact of model notations on the productivity of domain modelling: An empirical study. *Information and Software Technology*, 108, 78-87. doi:<https://doi.org/10.1016/j.infsof.2018.12.005>
- Farias, K., Garcia, A., Whittle, J., von Flach Garcia Chavez, C., & Lucena, C. (2015). Evaluating the effort of composing design models: a controlled experiment. *Software & Systems Modeling*, 14(4), 1349-1365. doi:10.1007/s10270-014-0408-2
- Frank, U. (2014). Multilevel Modeling. *Business & Information Systems Engineering*, 6(6), 319-337. doi:10.1007/s12599-014-0350-4
- Gemino, A., & Wand, Y. (2004). A framework for empirical evaluation of conceptual modeling techniques. *Requirements Engineering*, 9(4), 248-260. doi:10.1007/s00766-004-0204-6
- Jošt, G., Huber, J., Heričko, M., & Polančič, G. (2016). An empirical investigation of intuitive understandability of process diagrams. *Computer Standards & Interfaces*, 48, 90-111. doi:<https://doi.org/10.1016/j.csi.2016.04.006>
- Pollaert, W. (2016). Modelleren Voor Requirements. Retrieved from <https://www.vanharen.net/blog/modelleren-voor-requirements-in-3-minuten-2/>
- Roelens, B., & Bork, D. (2020). Performing an experimental evaluation to improve the initial PGA notation.
- Roelens, B., & Poels, G. (2015). The Development and Experimental Evaluation of a Focused Business Model Representation. *Business & Information Systems Engineering*, 57(1), 61-71. doi:10.1007/s12599-014-0363-z
- Roelens, B., Steenacker, W., & Poels, G. (2019). Realizing strategic fit within the business architecture: the design of a Process-Goal Alignment modeling and analysis technique. *Software & Systems Modeling*, 18(1), 631-662. doi:10.1007/s10270-016-0574-5
- Saunders, M. N. K. L., Philip; Thornhill, Adrian. (2019). *Research Methods for Business Students - eight edition* (Vol. eight edition). Harrow, UK: Pearson Education Limited.
- Tonella, P., Torchiano, M., Penta, M., Ceccato, M., & Ricca, F. (2010). How Developers' Experience and Ability Influence Web Application Comprehension Tasks Supported by UML Stereotypes: A Series of Four Experiments. *IEEE Transactions on Software Engineering*, 36(1), 96-118. doi:10.1109/TSE.2009.69
- Trkman, M., Mendling, J., Trkman, P., & Krisper, M. (2019). Impact of the conceptual model's representation format on identifying and understanding user stories. *Information and Software Technology*, 116, 106169. doi:<https://doi.org/10.1016/j.infsof.2019.08.001>

Bijlage 1: Notatie Associatie

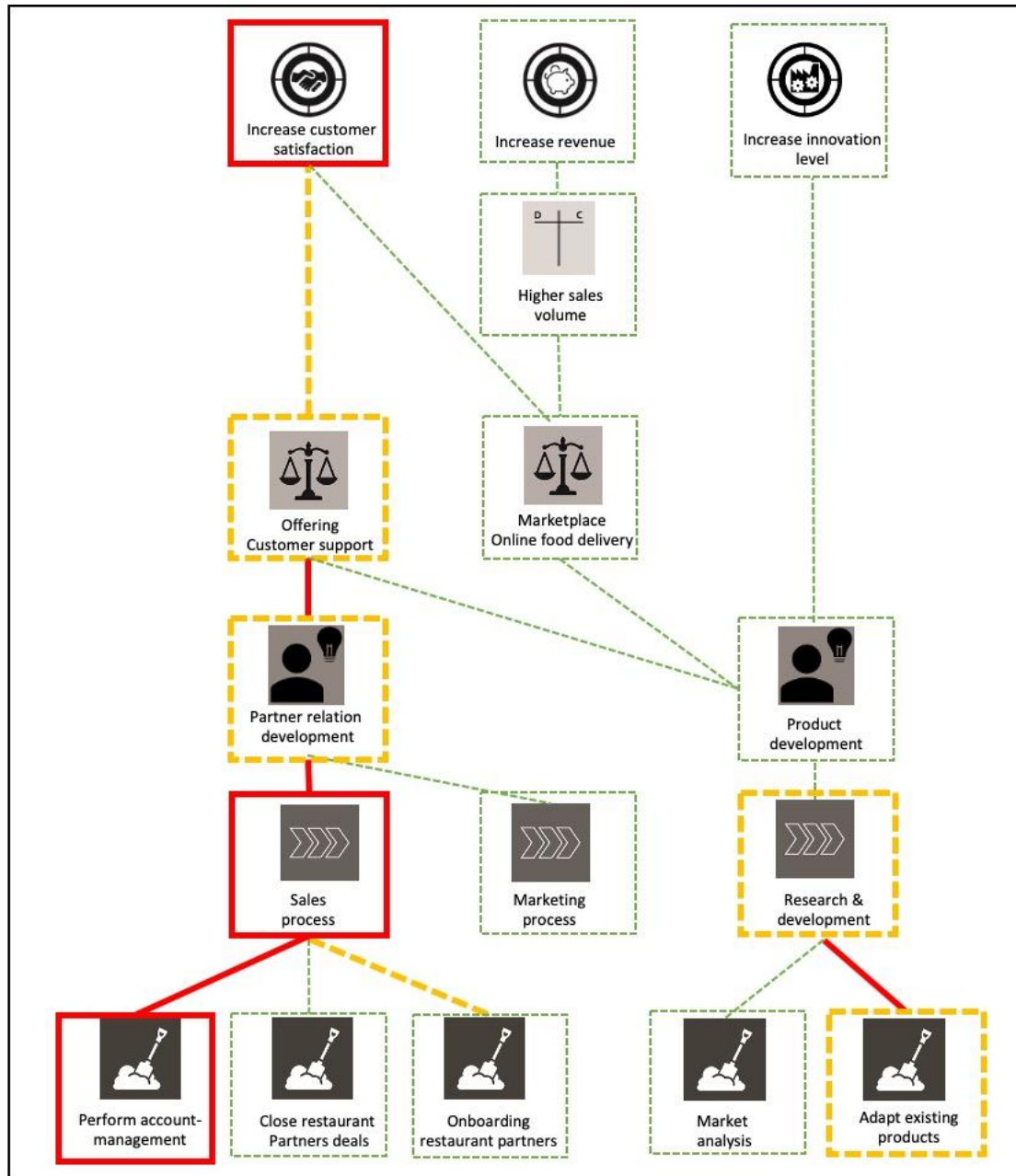


Scoringsmodel:

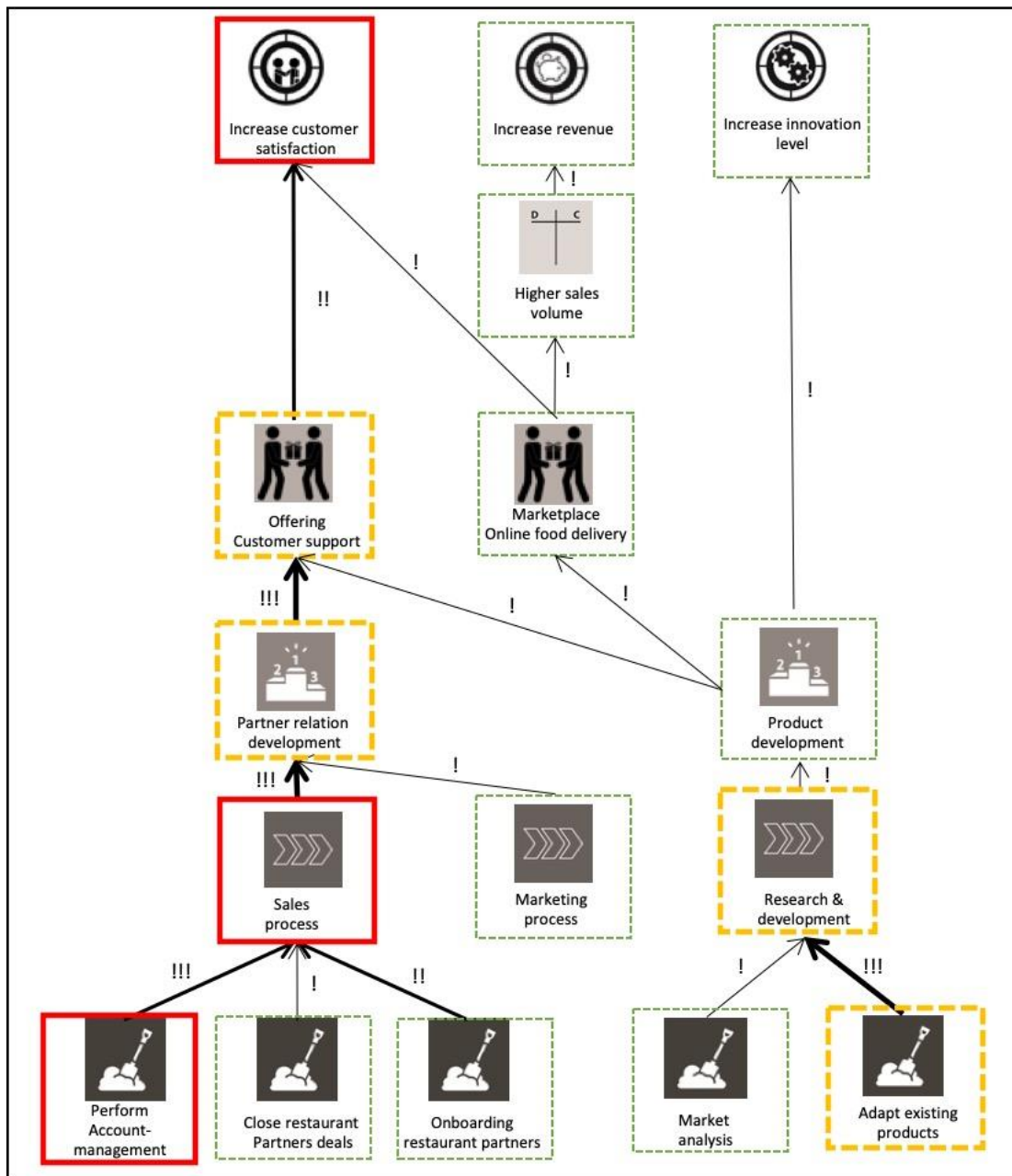
Ander meta model element	-1
Fout	0
Gedeeltelijk	1
Goed	2

Bijlage 2: Casus voor Comprehension Questions

Casus Groep A



Casus Groep B








Scoringsmodel






Fout	0
Goed	1

Bijlage 3: Tabellen Recallquestions

Tabel Groep A

Notation	Concept	Definition
	Internal Goal	Strategic objective that describes a desired state or development of the company to ensure that a company excels in processes, decisions, and actions to meet its customers' expectations.
	Value Proposition	Offered set of products and/or services that provides value to the customers and other partners, and competes in the overall value network.
	Competence	An integrated and holistic set of knowledge, skills, and abilities, related to a specific set of resources, which is coordinated through processes to realize the intended value proposition.
	Value stream	Representation of the hierarchical structure, through which value is created at distinct levels in the business architecture. Value streams can be differentiated based on three levels of strategic importance: high, medium, or low.
	Customer Goal	Strategic objective that describes a desired state or development of the company to ensure that a customer is satisfied with the products & services the company provides.

Tabel Groep B

Notation	Concept	Definition
	Internal Goal	Strategic objective that describes a desired state or development of the company to ensure that a company excels in processes, decisions, and actions to meet its customers' expectations.
	Value Proposition	Offered set of products and/or services that provides value to the customers and other partners, and competes in the overall value network.
	Competence	An integrated and holistic set of knowledge, skills, and abilities, related to a specific set of resources, which is coordinated through processes to realize the intended value proposition.
	Value stream	Representation of the hierarchical structure, through which value is created at distinct levels in the business architecture. Value streams can be differentiated based on three levels of strategic importance: high, medium, or low.
	Customer Goal	Strategic objective that describes a desired state or development of the company to ensure that a customer is satisfied with the products & services the company provides.

Ander meta model element	-1
Fout	0
Gedeeltelijk	1
Goed	2

Bijlage 4: Statistische testen individuele dataset

Notatie Associatie

Normaliteitstest

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
AV: Notation Association - Q1	.511	15	.000	.430	15	.000
AV: Notation Association - Q2	.	15	.	.	15	.
AV: Notation Association - Q3	.	15	.	.	15	.
AV: Notation Association - Q4	.428	15	.000	.661	15	.000
AV: Notation Association - Q5	.535	15	.000	.284	15	.000

a. Lilliefors Significance Correction

Uit de normaliteitstest blijkt dat de significantie in alle gevallen klein is dan 0,05 of in het geheel niet bestaand.

Hiermee wordt de nul-hypothese dat er sprake is van een normale verdeling voor alle notatie-associaties verworpen.

Dit betekent dat er sprake is van een intervalverdeling. Daarnaast is er sprake van twee groepen en van niet-gepaarde data. Om nu te kunnen bepalen of er tussen de groepen verschil bestaat in de associatie tussen oude of de nieuwe notatie kijken we naar onderstaande tabel.

Type of dependent variable	Type of independent variable						
	Ordinal/categorical				Normal/interval (ordinal)	More than 1	None
	Two groups		More groups				
	Paired	Unpaired	Paired	Unpaired			
2 categories	McNemar Test, Sign-Test	Fisher Test, Chi-squared-Test	Cochran's Q-Test	Fisher Test, Chi-squared-test	(Conditional) Logistic Regression	Logistic Regression	Chi-squared-Test
Nominal	Bowker Test	Fisher Test, Chi-squared-Test		Fisher Test, Chi-squared-test	Multinomial logistic regression	Multinomial logistic regression	Binomial Test
Ordinal	Wilcoxon Test, Sign-Test	Wilcoxon-Mann-Whitney Test	Friedman-Test	Kruskal-Wallis Test	Spearman-rank-test	Ordered logit	Median Test
Interval	Wilcoxon Test, Sign-Test	Wilcoxon-Mann-Whitney Test	Friedman-Test	Kruskal-Wallis Test	Spearman-rank test	Multivariate linear model	Median Test
Normal	t-Test (for paired)	t-Test (for unpaired)	Linear Model (ANOVA)	Linear Model (ANOVA)	Pearson-Correlation-test	Multivariate Linear Model	t-Test
Censored Interval	Log-Rank Test		Survival Analysis, Cox proportional hazards regression				
None	Clustering, factor analysis, PCA, canonical correlation						

Wilcoxon - Mann-Whitney Test

Ranks				
	VersionCode	N	Mean Rank	Sum of Ranks
AV: Notation Association - Q1	0	10	7.50	75.00
	1	5	9.00	45.00
	Total	15		
AV: Notation Association - Q2	0	10	8.00	80.00
	1	5	8.00	40.00
	Total	15		
AV: Notation Association - Q3	0	10	8.00	80.00
	1	5	8.00	40.00
	Total	15		
AV: Notation Association - Q4	0	10	9.60	96.00
	1	5	4.80	24.00
	Total	15		
AV: Notation Association - Q5	0	10	8.25	82.50
	1	5	7.50	37.50
	Total	15		

Test Statistics ^a					
	AV: Notation Association - Q1	AV: Notation Association - Q2	AV: Notation Association - Q3	AV: Notation Association - Q4	AV: Notation Association - Q5
Mann-Whitney U	20.000	25.000	25.000	9.000	22.500
Wilcoxon W	75.000	40.000	40.000	24.000	37.500
Z	-1.035	.000	.000	-2.515	-.707
Asymp. Sig. (2-tailed)	.301	1.000	1.000	.012	.480
Exact Sig. [2*(1-tailed Sig.)]	.594 ^b	1.000 ^b	1.000 ^b	.055 ^b	.768 ^b

a. Grouping Variable: VersionCode

b. Not corrected for ties.

- H0 Notatie associatie: Score PGA-1 = Score PGA-2
- H1 Notatie associatie: Score PGA-2 > Score PGA-1

Q1: Internal goal:

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.301 → 1-tailed sig = $(1 - (0.301 / 2)) = 0.8495 > 0,05$ H0 → kan niet verworpen worden.

Q2: Value Proposition

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 = PGA-1

2-tailed Sig = 1.000 → 1-tailed sig = $1.000 / 2 = 0.5000 > 0,05$ H0 → kan niet verworpen worden

Q3: Competence

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Score PGA-2 = PGA-1

2-tailed Sig = 1.000 → 1-tailed sig = $(1 - (1.000 / 2)) = 0.5000 > 0,05$ H0 → kan niet verworpen worden

Q4: Value Stream

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.012 → 1-tailed sig = $(1 - (0,012 / 2)) = 0.994 > 0,05$ H0 → kan niet verworpen worden

Q5: Customer Goal

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.480 → 1-tailed sig = $(1 - (0.480/2)) = 0.240 > 0,760$ H0 → kan niet verworpen worden

Conclusie:

Hoewel er uit de data blijkt dat er verschillen bestaan tussen de associatie van PGA-1 en PGA-2 is significantie van deze verschillen niet aan te tonen. Voorafgaand aan het onderzoek was geen enkele van de respondenten bekend met de PGA-taal.

Vanuit de data lijken de resultaten erop te wijzen dat een (gedeeltelijk) juiste intuïtieve associatie lijkt te berusten op toeval. De dataset is beperkt < 30. Dit is mogelijk van invloed op de resultaten.

Toetsing aan de gecombineerde dataset van alle onderzoekers kan mogelijk meer uitsluitsel bieden.

Notatie Associatie – Group time

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Group time: Notation association - Q1	.299	15	.001	.605	15	.000
Group time: Notation association - Q2	.397	15	.000	.407	15	.000
Group time: Notation association - Q3	.279	15	.003	.647	15	.000
Group time: Notation association - Q4	.124	15	.200 [*]	.952	15	.561
Group time: Notation association - Q5	.308	15	.000	.745	15	.001

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Uit de normaliteitstest blijkt dat de significantie in alle gevallen kleiner is dan 0,05, met uitzondering van de significantie voor Q4

Hiermee wordt de nul-hypothese dat er sprake is van een normale verdeling voor de group time Q1 – Q3 en Q5 verworpen.

Voor Q4 wordt de 0 hypothese niet verworpen.

Dit betekent dat er sprake is van een intervalverdeling voor Q1 – q3 en Q5 en een normale verdeling voor Q4. Daarnaast is er sprake van twee groepen en van niet-gepaarde data.

Voor de vragen Q1-Q3 en Q5 wordt de Wilcoxon-Mann-Whitneytest uitgevoerd, voor Q4 de t-Test

Type of dependent variable	Type of independent variable						
	Ordinal/categorical				Normal/interval (ordinal)	More than 1	None
	Two groups		More groups				
	Paired	Unpaired	Paired	Unpaired			
2 categories	McNemar Test, Sign-Test	Fisher Test, Chi-squared-Test	Cochran's Q-Test	Fisher Test, Chi-squared-test	(Conditional) Logistic Regression	Logistic Regression	Chi-squared-Test
Nominal	Bowker Test	Fisher Test, Chi-squared-Test		Fisher Test, Chi-squared-test	Multinomial logistic regression	Multinomial logistic regression	Binomial Test
Ordinal	Wilcoxon Test, Sign-Test	Wilcoxon-Mann-Whitney Test	Friedman-Test	Kruskal-Wallis Test	Spearman-rank-test	Ordered logit	Median Test
Interval	Wilcoxon Test, Sign-Test	Wilcoxon-Mann-Whitney Test	Friedman-Test	Kruskal-Wallis Test	Spearman-rank test	Multivariate linear model	Median Test
Normal	t-Test (for paired)	t-Test (for unpaired)	Linear Model (ANOVA)	Linear Model (ANOVA)	Pearson-Correlation-test	Multivariate Linear Model	t-Test
Censored Interval	Log-Rank Test		Survival Analysis, Cox proportional hazards regression				
None	Clustering, factor analysis, PCA, canonical correlation						

Wilcoxon - Mann-Whitney Test

Ranks

	VersionCode	N	Mean Rank	Sum of Ranks
Group time: Notation association - Q1	0	10	7.50	75.00
	1	5	9.00	45.00
	Total	15		
Group time: Notation association - Q2	0	10	7.40	74.00
	1	5	9.20	46.00
	Total	15		
Group time: Notation association - Q3	0	10	8.30	83.00
	1	5	7.40	37.00
	Total	15		
Group time: Notation association - Q5	0	10	7.70	77.00
	1	5	8.60	43.00
	Total	15		

Test Statistics^a

	Group time: Notation association - Q1	Group time: Notation association - Q2	Group time: Notation association - Q3	Group time: Notation association - Q5
Mann-Whitney U	20.000	19.000	22.000	22.000
Wilcoxon W	75.000	74.000	37.000	77.000
Z	-.612	-.735	-.367	-.367
Asymp. Sig. (2-tailed)	.540	.462	.713	.713
Exact Sig. [2*(1-tailed Sig.)]	.594 ^b	.513 ^b	.768 ^b	.768 ^b

a. Grouping Variable: VersionCode

b. Not corrected for ties.

- H0 Notatie associatie: Grouptime PGA-1 snelheid = Grouptime PGA-2 snelheid
- H1 Notatie associatie: Grouptime PGA-2 snelheid < Grouptime PGA-1 snelheid

Q1: Grouptime Internal goal:

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.540 → 1-tailed sig = $0.540 / 2 = 0.270 > 0,05$ H0 → kan niet verworpen worden

Q2: Grouptime Value Proposition

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 > PGA-1

2-tailed Sig = 0.462 → 1-tailed sig = $(1 - (0.462 / 2)) = 0.769 > 0,05$ H0 → kan niet verworpen worden

Q3: Competence

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Score PGA-2 > PGA-1

2-tailed Sig = 0.713 → 1-tailed sig = $(1 - (0.713 / 2)) = 0.6435 > 0,05$ H0 → kan niet verworpen worden

Q5: Customer Goal

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 > PGA-1

2-tailed Sig = 0.713 → 1-tailed sig = $(1 - (0.713 / 2)) = 0.6435 > 0,05$ H0 → kan niet verworpen worden

T-Test

Group Statistics					
	VersionCode	N	Mean	Std. Deviation	Std. Error Mean
Group time: Notation association - Q4	0	10	45.9520000000 00000	20.3513629355 21870	6.43566603649 7957
	1	5	31.1680000000 00000	22.2091134447 10034	9.93221747647 5230

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Group time: Notation association - Q4	Equal variances assumed	.163	.693	1.289	13	.220	14.7839999 99999997	11.4696059 61979553	- 9.99457722 2239310	39.56257722 2239310
	Equal variances not assumed			1.249	7.478	.249	14.7839999 99999997	11.8349795 66240632	- 12.8426721 62641207	42.41067216 2641205

Q4: Value Stream

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Mean PGA-2 < PGA-1

Levene's test: sig.0693 > 0.05 Dit betekent dat er aan de aanname van homogene varianties is voldaan voor het symbool voor Value Stream

2-tailed Sig = 0.220 → 1-tailed sig = 0.220/2 = 0.110 > 0,05 H0 → kan niet verworpen worden

Conclusie:

Hoewel er uit de data blijkt dat er verschillen blijken bestaan tussen de gebruikte tijd voor PGA-1 en PGA-2 is significantie van deze verschillen niet aan te tonen. Voorafgaand aan het onderzoek was geen enkele van de respondenten bekend met de PGA-taal. Met andere woorden. Er kan niet aangetoond worden dat PGA-2 leidt tot een snellere (gedeeltelijk) juiste associatie.

De dataset is echter relatief klein (<30), mogelijk biedt de gecombineerde dataset van alle onderzoekers een voldoende grote dataset om meer eenduidige conclusies te kunnen trekken.

Comprehension Questions

Uitgaan van niet-normale verdeling. Daarbij uitvoeren Wilcoxon-Mann-Whitneytest

Ranks				
	VersionCode	N	Mean Rank	Sum of Ranks
CQQ1Code	0	10	8.00	80.00
	1	5	8.00	40.00
	Total	15		
CQQ2Code	0	10	9.00	90.00
	1	5	6.00	30.00
	Total	15		
CQQ3Code	0	10	7.75	77.50
	1	5	8.50	42.50
	Total	15		
CQQ4Code	0	10	8.00	80.00
	1	5	8.00	40.00
	Total	15		
CQQ5Code	0	10	7.25	72.50
	1	5	9.50	47.50
	Total	15		
CQQ6Code	0	10	8.50	85.00
	1	5	7.00	35.00
	Total	15		

Test Statistics ^a						
	CQQ1Code	CQQ2Code	CQQ3Code	CQQ4Code	CQQ5Code	CQQ6Code
Mann-Whitney U	25.000	15.000	22.500	25.000	17.500	20.000
Wilcoxon W	40.000	30.000	77.500	40.000	72.500	35.000
Z	.000	-1.414	-.707	.000	-1.323	-.798
Asymp. Sig. (2-tailed)	1.000	.157	.480	1.000	.186	.425
Exact Sig. [2*(1-tailed Sig.)]	1.000 ^b	.254 ^b	.768 ^b	1.000 ^b	.371 ^b	.594 ^b

a. Grouping Variable: VersionCode

b. Not corrected for ties.

- H0 Comprehension Questions: Score PGA-1 = Score PGA-2
- H1 Comprehension Questions: Score PGA-2 > Score PGA-1

Q1: Customer Goal

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 = PGA-1

2-tailed Sig = 1.000 → 1-tailed sig = $(1 - (1.000 / 2)) = 0.5000 > 0,05$ H0 → kan niet verworpen worden

Q2 Value Proposition

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.157 → 1-tailed sig = $(1 - (0.157 / 2)) = 0.9215 > 0,05$ H0 → kan niet verworpen worden

Q3: Value Stream

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 > PGA-1

2-tailed Sig = 0.480 → 1-tailed sig = $0,480 / 2 = 0.240 > 0,05$ H0 → kan niet verworpen worden

Q4: Competence

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Score PGA-2 = PGA-1

2-tailed Sig = 1.000 → 1-tailed sig = $1.000 / 2 = 0.5000 > 0,05$ H0 → kan niet verworpen worden

Q5: Internal goal:

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.186 → 1-tailed sig = $(1 - (0.186 / 2)) = 0.907 > 0,05$ H0 → kan niet verworpen worden

Q6: Importance *

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.425 → 1-tailed sig = $(1 - (0.425 / 2)) = 0.7875 > 0,05$ H0 → kan niet verworpen worden

*de case study is de enige opdracht waarin dit element werd gevraagd.

Conclusie:

Hoewel er uit de data blijkt dat er verschillen bestaan in het oplossen van de casestudy tussen PGA-1 en PGA-2 is significantie van deze verschillen niet aan te tonen. Voorafgaand aan de taak hebben de respondenten een summiere uitleg van de PGA taal gehad.

Comprehension Questions Group time

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Group time: Comprehension questions - Q1	.302	15	.001	.734	15	.001
Group time: Comprehension questions - Q2	.278	15	.003	.674	15	.000
Group time: Comprehension questions - Q3	.177	15	.200*	.882	15	.051
Group time: Comprehension questions - Q4	.186	15	.174	.880	15	.048
Group time: Comprehension questions - Q5	.361	15	.000	.682	15	.000
Group time: Comprehension questions - Q6	.238	15	.022	.753	15	.001

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Uit de normaliteitstest blijkt dat de significantie in alle gevallen kleiner is dan 0,05 , met uitzondering van de significantie voor Q3. Daarnaast is er sprake van twee groepen en van niet-gepaarde data.

Dit betekent een Wilcoxon-Mann-Whitney test voor de vragen Q1 en Q2 en Q4 tot en met Q6 en een t-Test voor niet-gepaarde data voor Q3

- H0 Comprehension: Grouptime PGA-1 snelheid = Grouptime PGA-2 snelheid
- H1 Comprehension: Grouptime PGA-2 snelheid < Grouptime PGA-1 snelheid

Wilcoxon-Mann-Whitney Test

Ranks				
	VersionCode	N	Mean Rank	Sum of Ranks
Group time: Comprehension questions - Q1	0	10	7.30	73.00
	1	5	9.40	47.00
	Total	15		
Group time: Comprehension questions - Q2	0	10	8.90	89.00
	1	5	6.20	31.00
	Total	15		
Group time: Comprehension questions - Q4	0	10	7.70	77.00
	1	5	8.60	43.00
	Total	15		
Group time: Comprehension questions - Q5	0	10	9.00	90.00
	1	5	6.00	30.00
	Total	15		
Group time: Comprehension questions - Q6	0	10	10.50	105.00
	1	5	3.00	15.00
	Total	15		

Q1: Customer Goal

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 > PGA-1

2-tailed Sig = 0.391 → 1-tailed sig = $(1 - (0.391 / 2)) = 0.8045 > 0,05$ H0 → kan niet verworpen worden

Q2 Value Proposition

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.270 → 1-tailed sig = $0.270 / 2 = 0.135 > 0,05$ H0 → kan niet verworpen worden

Q4: Competence

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.713 → 1-tailed sig = $0.713 / 2 = 0.3565 > 0,05$ H0 → kan niet verworpen worden

Q5: Internal goal:

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Score PGA-2 > PGA-1

2-tailed Sig = 0.221 → 1-tailed sig = $(1 - (0.221 / 2)) = 0.8895 > 0,05$ H0 → kan niet verworpen worden

Q6: Importance *

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.022 → 1-tailed sig = $0.02 / 2 = 0.01 < 0,05$ H1 → **kan aanvaard worden**

t-Test independent samples

Group Statistics					
	VersionCode	N	Mean	Std. Deviation	Std. Error Mean
Group time:	0	10	38.4610	9.09939	2.87748
Comprehension	1	5	41.9620	24.94302	11.15486
questions - Q3					

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Group time: Comprehension questions - Q3	Equal variances assumed	3.117	.101	-.405	13	.692	-3.50100	8.63865	-22.16368	15.16168
	Equal variances not assumed			-.304	4.541	.775	-3.50100	11.52002	-34.03725	27.03525

Q3: Value Stream

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 > PGA-1

Levene's test: sig = 0,101 > 0,05. Hiermee kan H₀, er is sprake van gelijke varianties, worden aanvaard.

2-tailed Sig = 0,692 → 1-tailed sig = $(1 - (0.692 / 2)) = 0.654 > 0,05$ H₀ → kan niet verworpen worden

Conclusie:

Hoewel er uit de data blijkt dat er verschillen bestaan in de besteedde tijd aan het oplossen van de casestudy tussen PGA-1 en PGA-2 is significantie van deze verschillen niet aan te tonen, met uitzondering van Q6. Hierbij zijn de verschillen tussen de groepen significant, waarmee gesteld kan worden dat voor Q6 met zekerheid geldt dat PGA-2 leidt tot een snellere oplossing van de taak. Voorafgaand aan de taak hebben de respondenten een summiere uitleg van de PGA taal gehad.

Recall Questions

Ranks				
	VersionCode	N	Mean Rank	Sum of Ranks
RQQ1Code	0	10	9.15	91.50
	1	5	5.70	28.50
	Total	15		
RQQ2	0	10	9.05	90.50
	1	5	5.90	29.50
	Total	15		
RQQ3Code	0	10	8.50	85.00
	1	5	7.00	35.00
	Total	15		
RQQ4Code	0	10	9.75	97.50
	1	5	4.50	22.50
	Total	15		
RQQ5Code	0	10	8.35	83.50
	1	5	7.30	36.50
	Total	15		

Test Statistics ^a					
	RQQ1Code	RQQ2	RQQ3Code	RQQ4Code	RQQ5Code
Mann-Whitney U	13.500	14.500	20.000	7.500	21.500
Wilcoxon W	28.500	29.500	35.000	22.500	36.500
Z	-1.506	-1.375	-.707	-2.565	-.483
Asymp. Sig. (2-tailed)	.132	.169	.480	.010	.629
Exact Sig. [2*(1-tailed Sig.)]	.165 ^b	.206 ^b	.594 ^b	.028 ^b	.679 ^b

a. Grouping Variable: VersionCode

b. Not corrected for ties.

- H0 Recall: Score PGA-1 = Score PGA-2
- H1 Recall: Score PGA-2 > Score PGA-1

Q1: Internal goal:

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Score PGA-2 > PGA-1

2-tailed Sig = 0.132 → 1-tailed sig = 0,132 / 2 = 0.066 > 0,05 H0 → kan niet verworpen worden

Q2: Value Proposition

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.132 → 1-tailed sig = $(1 - (0,132 / 2)) = 0.934 > 0,05$ H0 → kan niet verworpen worden

Q3: Competence

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Score PGA-2 > PGA-1

2-tailed Sig = 0.480 → 1-tailed sig = $0.480 / 2 = 0.240 > 0,05$ H0 → kan niet verworpen worden

Q4: Value Stream

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.010 → 1-tailed sig = $(1 - (0,010 / 2)) = 0.995 > 0,05$ H0 → kan niet verworpen worden

Q5: Customer Goal

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.629 → 1-tailed sig = $(1 - (0.629/2)) = 0.6855 > 0,05$ H0 → kan niet verworpen worden

Conclusie:

Hoewel er uit de data blijkt dat er verschillen bestaan in het herinneren van PGA-1 en PGA-2 is significantie van deze verschillen niet aan te tonen. In geval van de symbolen voor Competence en Customer Goal lijkt het zo te zijn dat deze beter herinnerd worden. Voorafgaand aan de taak hebben de respondenten gedurende een minuut de symbolen en hun betekenis in zich op kunnen nemen.

Recall Questions Group Time

	Tests of Normality					
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Group time: Recall questions - Q1	.265	15	.006	.869	15	.033
Group time: Recall questions - Q2	.183	15	.188	.907	15	.123
Group time: Recall questions - Q3	.229	15	.034	.789	15	.003
Group time: Recall questions - Q4	.242	15	.018	.772	15	.002
Group time: Recall questions - Q5	.096	15	.200 [*]	.985	15	.992

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Q1, Q3 en Q4 hebben een sig. < 0.05 oftewel een niet-normale verdeling

Q2 en Q5 hebben een sig > 0.05 oftewel een normale verdeling

Daarnaast is er sprake van twee groepen en van niet-gepaarde data.

Dit betekent een Wilcoxon-Mann-Whitney test voor Q1, Q3 en Q4 en een t-test voor ongepaarde data voor Q2 en Q5

- H0 Recall: Grouptime PGA-1 snelheid = Grouptime PGA-2 snelheid
- H1 Recall: Grouptime PGA-2 snelheid < Grouptime PGA-1 snelheid

Wilcoxon-Mann-Whitney Test

Ranks				
	VersionCode	N	Mean Rank	Sum of Ranks
Group time: Recall questions	0	10	8.60	86.00
- Q1	1	5	6.80	34.00
	Total	15		
Group time: Recall questions	0	10	8.20	82.00
- Q3	1	5	7.60	38.00
	Total	15		
Group time: Recall questions	0	10	6.90	69.00
- Q4	1	5	10.20	51.00
	Total	15		

Test Statistics ^a			
	Group time: Recall questions - Q1	Group time: Recall questions - Q3	Group time: Recall questions - Q4
Mann-Whitney U	19.000	23.000	14.000
Wilcoxon W	34.000	38.000	69.000
Z	-.735	-.245	-1.347
Asymp. Sig. (2-tailed)	.462	.806	.178
Exact Sig. [2*(1-tailed Sig.)]	.513 ^b	.859 ^b	.206 ^b

a. Grouping Variable: VersionCode

b. Not corrected for ties.

Q1: Internal goal:

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Ranking: Score PGA-2 > PGA-1

2-tailed Sig = 0.462 → 1-tailed sig = $(1 - (0,462 / 2)) = 0.769 > 0,05 H_0 \rightarrow$ kan niet verworpen worden

Q3: Competence

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Score PGA-2 > PGA-1

2-tailed Sig = 0.806 → 1-tailed sig = $(1 - (0,806 / 2)) = 0.597 > 0,05 H_0 \rightarrow$ kan niet verworpen worden

Q4: Value Stream

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 > PGA-1

2-tailed Sig = 0.178 → 1-tailed sig = $(1 - (0,178 / 2)) = 0.911 > 0,05$ H0 → kan niet verworpen worden

t-Test independent samples

Group Statistics

	VersionCode	N	Mean	Std. Deviation	Std. Error Mean
Group time: Recall questions - Q2	0	10	14.062000000000001	9.327225144108450	2.949527570457494
	1	5	14.014000000000000	5.929264709894472	2.651647789582922
Group time: Recall questions - Q5	0	10	11.362999999999998	3.841487125233317	1.214784891794977
	1	5	10.596000000000000	2.413261278850676	1.079243253395637

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
Group time: Recall questions - Q2	Equal variances assumed	.728	.409	.010	13	.992	.0480000000	4.616689323	-9.92575091	10.02175091
	Equal variances not assumed			.012	11.915	.991	.0480000000	3.966226026	-8.60050672	8.696506720
Group time: Recall questions - Q5	Equal variances assumed	2.113	.170	.404	13	.693	.7670000000	1.898026485	-3.33343693	4.867436928
	Equal variances not assumed			.472	11.997	.645	.7670000000	1.624951794	-2.77355385	4.307553847

Q2: Value Proposition

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Ranking: Score PGA-2 < PGA-1

Levene's test: sig = 0,409 > 0,05. Hiermee kan H0, er is sprake van gelijke varianties, worden aanvaard.

2-tailed Sig = 0.992 → 1-tailed sig = 0,992 / 2 = 0.492 > 0,05 H0 → kan niet verworpen worden

Q5: Customer Goal

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 < PGA-1

Levene's test: sig = 0.170 > 0,05. Hiermee kan H0, er is sprake van gelijke varianties, worden aanvaard.

2-tailed Sig = 0.806 → 1-tailed sig = 0.693 / 2 = 0.3465 > 0,05 H0 → kan niet verworpen worden

Conclusie:

Uit de data blijkt dat, met uitzondering van de group time voor de symbolen voor Value Proposition en Customer Goal, de tijd om PGA-2 te herinneren langer is dan de tijd om PGA-1 te herinneren.

De verschillen blijken echter niet significant. Kortere tijd nodig om te herinneren duidt niet per se op meer intuïtief zijn van de notatie.

Omdat de dataset van de individuele onderzoeker relatief klein is (< 30), is het belangrijk om deze conclusie te toetsen aan de grotere, gecombineerde dataset (> 30) van alle drie de onderzoekers.

Representatievoorkeur

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
CustomerGoalCode	.124	15	.200*	.935	15	.324
ValuePropositionCode	.157	15	.200*	.899	15	.091
CompetenceCode	.178	15	.200*	.907	15	.121
InternalGoalCode	.129	15	.200*	.963	15	.746
ValueStreamCode	.166	15	.200*	.903	15	.106

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Uit de normaliteitstest blijkt dat de significantie in alle gevallen groter is dan 0,05

Hiermee wordt de nul-hypothese dat er sprake is van een normale verdeling voor alle representatievoorkeuren niet verworpen.

Dit betekent dat er sprake is van een normale verdeling. Daarnaast is er sprake van twee groepen en van gepaarde data.

Gebruikmakend van onderstaande tabel volgt hieruit een t-Test voor gepaarde data.

Type of dependent variable	Type of independent variable						
	Ordinal/categorical				Normal/interval (ordinal)	More than 1	None
	Two groups		More groups				
	Paired	Unpaired	Paired	Unpaired			
2 categories	McNemar Test, Sign-Test	Fisher Test, Chi-squared-Test	Cochran's Q-Test	Fisher Test, Chi-squared-test	(Conditional) Logistic Regression	Logistic Regression	Chi-squared-Test
Nominal	Bowker Test	Fisher Test, Chi-squared-Test		Fisher Test, Chi-squared-test	Multinomial logistic regression	Multinomial logistic regression	Binomial Test
Ordinal	Wilcoxon Test, Sign-Test	Wilcoxon-Mann-Whitney Test	Friedman-Test	Kruskal-Wallis Test	Spearman-rank-test	Ordered logit	Median Test
Interval	Wilcoxon Test, Sign-Test	Wilcoxon-Mann-Whitney Test	Friedman-Test	Kruskal-Wallis Test	Spearman-rank test	Multivariate linear model	Median Test
Normal	t-Test (for paired)	t-Test (for unpaired)	Linear Model (ANOVA)	Linear Model (ANOVA)	Pearson-Correlation-test	Multivariate Linear Model	t-Test
Censored Interval	Log-Rank Test		Survival Analysis, Cox proportional hazards regression				
None	Clustering, factor analysis, PCA, canonical correlation						

H0: Er is een voorkeur voor de nieuwe notatie

H1: Er is geen voorkeur voor de oude of de nieuwe notatie

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
CustomerGoalCode	15	53.67	34.094	8.803
ValuePropositionCode	15	63.67	33.672	8.694
CompetenceCode	15	40.00	32.071	8.281
InternalGoalCode	15	54.67	29.909	7.722
ValueStreamCode	15	55.00	35.203	9.090

One-Sample Test

				Test Value = 50		
				Mean Difference	95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)		Lower	Upper
CustomerGoalCode	.417	14	.683	3.667	-15.21	22.55
ValuePropositionCode	1.572	14	.138	13.667	-4.98	32.31
CompetenceCode	-1.208	14	.247	-10.000	-27.76	7.76
InternalGoalCode	.604	14	.555	4.667	-11.90	21.23
ValueStreamCode	.550	14	.591	5.000	-14.50	24.50

Customer Goal

Mean: 53,67. Hieruit lijkt een lichte voorkeur te blijken voor de nieuwe notatie van het symbool voor Customer Goal.

2-tailed sig. = 0.138 → 1-tailed sig = $0.683/2 = 0.3415 > 0.05$

De score is niet significant t.o.v. het midden van de schaal.

De dataset is < 30. Om die reden is het goed om dit resultaat van de individuele onderzoeker te vergelijken met de gecombineerde dataset van alle onderzoekers

Value Proposition

Mean: 63,67. Hieruit lijkt een voorkeur te blijken voor de nieuwe notatie van het symbool voor Customer Goal.

2-tailed sig. = 0.138 → 1-tailed sig = $0.138/2 = 0.069 > 0.05$

De score is niet significant t.o.v. het midden van de schaal.

De dataset is < 30. Om die reden is het goed om dit resultaat van de individuele onderzoeker te vergelijken met de gecombineerde dataset van alle onderzoekers.

Competence

Mean: 40,00. Hieruit lijkt een voorkeur te blijken voor de oude notatie van het symbool voor Customer Goal.

2-tailed sig. = 0.247 → 1-tailed sig = $0.247/2 = 0.1235 > 0.05$

De score is niet significant t.o.v. het midden van de schaal.

De dataset is < 30. Om die reden is het goed om dit resultaat van de individuele onderzoeker te vergelijken met de gecombineerde dataset van alle onderzoekers.

Internal Goal

Mean: 54,67. Hieruit lijkt een lichte voorkeur te blijken voor de nieuwe notatie van het symbool voor Customer Goal.

2-tailed sig. = 0.555 → 1-tailed sig = $0.555/2 = 0.2775 > 0.05$

De score is niet significant t.o.v. het midden van de schaal.

De dataset is < 30. Om die reden is het goed om dit resultaat van de individuele onderzoeker te vergelijken met de gecombineerde dataset van alle onderzoekers.

Valuestream

Mean: 55,00. Hieruit lijkt een voorkeur te blijken voor de nieuwe notatie van het symbool voor Customer Goal.

2-tailed sig. = 0.591 → 1-tailed sig = $0.591/2 = 0.2955 > 0.05$

De score is niet significant t.o.v. het midden van de schaal.

De dataset is < 30. Om die reden is het goed om dit resultaat van de individuele onderzoeker te vergelijken met de gecombineerde dataset van alle onderzoekers.

Conclusie:

Over het algemeen lijkt er een voorkeur te bestaan voor de nieuwe notatie. Dit zou kunnen betekenen dat de nieuwe notatie als intuïtiever wordt ervaren, met uitzondering van het symbool voor competence. Hier ligt de voorkeur ten gunste van de oude notatie. De dataset is echter relatief klein (<30) waardoor vergelijk met een grotere dataset noodzakelijk is.

Bijlage 5: Statistische testen gecombineerde dataset

Notatie Associatie

Normaliteitstest

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
AV: Notation Association - Q1	.351	44	.000	.764	44	.000
AV: Notation Association - Q2	.537	44	.000	.276	44	.000
AV: Notation Association - Q3	.535	44	.000	.284	44	.000
AV: Notation Association - Q4	.387	44	.000	.719	44	.000
AV: Notation Association - Q5	.391	44	.000	.666	44	.000

a. Lilliefors Significance Correction

Uit de normaliteitstest blijkt dat de significantie in alle gevallen klein is dan 0,05 of in het geheel niet bestaand.

Hiermee wordt de nul-hypothese dat er sprake is van een normale verdeling voor alle notatie-associaties verworpen.

Dit betekent dat er sprake is van een intervalverdeling. Daarnaast is er sprake van twee groepen en van niet-gepaarde data. Om nu te kunnen bepalen of er tussen de groepen verschil bestaat in de associatie tussen oude of de nieuwe notatie kijken we naar onderstaande tabel.

Type of dependent variable	Type of independent variable						
	Ordinal/categorical				Normal/interval (ordinal)	More than 1	None
	Two groups		More groups				
	Paired	Unpaired	Paired	Unpaired			
2 categories	McNemar Test, Sign-Test	Fisher Test, Chi-squared-Test	Cochran's Q-Test	Fisher Test, Chi-squared-test	(Conditional) Logistic Regression	Logistic Regression	Chi-squared-Test
Nominal	Bowker Test	Fisher Test, Chi-squared-Test		Fisher Test, Chi-squared-test	Multinomial logistic regression	Multinomial logistic regression	Binomial Test
Ordinal	Wilcoxon Test, Sign-Test	Wilcoxon-Mann-Whitney Test	Friedman-Test	Kruskal-Wallis Test	Spearman-rank-test	Ordered logit	Median Test
Interval	Wilcoxon Test, Sign-Test	Wilcoxon-Mann-Whitney Test	Friedman-Test	Kruskal-Wallis Test	Spearman-rank test	Multivariate linear model	Median Test
Normal	t-Test (for paired)	t-Test (for unpaired)	Linear Model (ANOVA)	Linear Model (ANOVA)	Pearson-Correlation-test	Multivariate Linear Model	t-Test
Censored Interval	Log-Rank Test		Survival Analysis, Cox proportional hazards regression				
None	Clustering, factor analysis, PCA, canonical correlation						

Wilcoxon - Mann-Whitney Test

Ranks				
	VersionCode	N	Mean Rank	Sum of Ranks
AV: Notation Association - Q1	0	24	23.13	555.00
	1	20	21.75	435.00
	Total	44		
AV: Notation Association - Q2	0	24	21.00	504.00
	1	20	24.30	486.00
	Total	44		
AV: Notation Association - Q3	0	24	22.85	548.50
	1	20	22.08	441.50
	Total	44		
AV: Notation Association - Q4	0	24	24.23	581.50
	1	20	20.43	408.50
	Total	44		
AV: Notation Association - Q5	0	24	23.10	554.50
	1	20	21.78	435.50
	Total	44		

Test Statistics^a					
	AV: Notation Association - Q1	AV: Notation Association - Q2	AV: Notation Association - Q3	AV: Notation Association - Q4	AV: Notation Association - Q5
Mann-Whitney U	225.000	204.000	231.500	198.500	225.500
Wilcoxon W	435.000	504.000	441.500	408.500	435.500
Z	-.429	-1.943	-.459	-1.214	-.406
Asymp. Sig. (2-tailed)	.668	.052	.647	.225	.684

a. Grouping Variable: VersionCode

- H0 Notatie associatie: Score PGA-1 = Score PGA-2
- H1 Notatie associatie: Score PGA-2 > Score PGA-1

Q1: Internal goal:

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Score PGA-2 > PGA-1

2-tailed Sig =0.668 → 1-tailed sig = 0.668 / 2= 0.334 > 0,05 H0 → kan niet verworpen worden.

Q2: Value Proposition

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 > PGA-1

2-tailed Sig = 0.052 → 1-tailed sig = $0.052 / 2 = 0.026 < 0,05$ H1 → **kan worden aanvaard**

Q3: Competence

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Score PGA-2 > PGA-1

2-tailed Sig = 0.647 → 1-tailed sig = $0.647 / 2 = 0.3235 > 0,05$ H0 → kan niet verworpen worden

Q4: Value Stream

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.225 → 1-tailed sig = $(1 - (0,225 / 2)) = 0.8875 > 0,05$ H0 → kan niet verworpen worden

Q5: Customer Goal

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.684 → 1-tailed sig = $(1 - (0.684/2)) = 0.240 > 0,05$ H0 → kan niet verworpen worden

Conclusie:

Hoewel er uit de data blijkt dat er verschillen bestaan tussen de associatie van PGA-1 en PGA-2 is significantie van deze verschillen niet aan te tonen. Hierbij scoren de symbolen voor Internal Goal, Value Proposition en Competence beter in de nieuwe notatie. De symbolen voor Valuestream en Customer Goal scoren beter in de oude notatie.

Voorafgaand aan het onderzoek was geen enkele van de respondenten bekend met de PGA-taal.

Vanuit de data lijken de resultaten erop te wijzen dat een (gedeeltelijk) juiste intuïtieve associatie lijkt te berusten op toeval.

Notatie Associatie – Group time

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Group time: Notation association - Q1	.191	44	.000	.705	44	.000
Group time: Notation association - Q2	.290	44	.000	.515	44	.000
Group time: Notation association - Q3	.204	44	.000	.685	44	.000
Group time: Notation association - Q4	.249	44	.000	.638	44	.000
Group time: Notation association - Q5	.319	44	.000	.357	44	.000

a. Lilliefors Significance Correction

Uit de normaliteitstest blijkt dat de significantie in alle gevallen kleiner is dan 0,05

Hiermee wordt de nul-hypothese dat er sprake is van een normale verdeling voor de group time verworpen. Dit betekent dat er sprake is van een intervalverdeling en dat de Wilcoxon-Mann-Whitneytest wordt uitgevoerd.

Wilcoxon-Mann-Whitney Test

Ranks

	VersionCode	N	Mean Rank	Sum of Ranks
Group time: Notation association - Q1	0	24	21.92	526.00
	1	20	23.20	464.00
	Total	44		
Group time: Notation association - Q2	0	24	20.58	494.00
	1	20	24.80	496.00
	Total	44		
Group time: Notation association - Q3	0	24	22.75	546.00
	1	20	22.20	444.00
	Total	44		
Group time: Notation association - Q4	0	24	24.88	597.00
	1	20	19.65	393.00
	Total	44		
Group time: Notation association - Q5	0	24	20.17	484.00
	1	20	25.30	506.00
	Total	44		

Test Statistics ^a					
	Group time: Notation association - Q1	Group time: Notation association - Q2	Group time: Notation association - Q3	Group time: Notation association - Q4	Group time: Notation association - Q5
Mann-Whitney U	226.000	194.000	234.000	183.000	184.000
Wilcoxon W	526.000	494.000	444.000	393.000	484.000
Z	-.330	-1.084	-.141	-1.344	-1.320
Asymp. Sig. (2-tailed)	.741	.278	.888	.179	.187

a. Grouping Variable: VersionCode

- H0 Notatie associatie: Grouptime PGA-1 snelheid = Grouptime PGA-2 snelheid
- H1 Notatie associatie: Grouptime PGA-2 snelheid < Grouptime PGA-1 snelheid

Q1: Grouptime Internal goal:

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.741 → 1-tailed sig = $0.741 / 2 = 0.3705 > 0,05$ H0 → kan niet verworpen worden

Q2: Grouptime Value Proposition

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 > PGA-1

2-tailed Sig = 0.278 → 1-tailed sig = $(1 - (0.278 / 2)) = 0.861 > 0,05$ H0 → kan niet verworpen worden

Q3: Competence

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Score PGA-2 > PGA-1

2-tailed Sig = 0.888 → 1-tailed sig = $(1 - (0.888 / 2)) = 0.556 > 0,05$ H0 → kan niet verworpen worden

Q4: Value Stream

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Mean PGA-2 < PGA-1

2-tailed Sig = 0.179 → 1-tailed sig = $0.179/2 = 0.0895 > 0,05$ H0 → kan niet verworpen worden

Q5: Customer Goal

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 > PGA-1

2-tailed Sig = 0.187 → 1-tailed sig = $(1 - (0.187 / 2)) = 0.9065 > 0,05$ H0 → kan niet verworpen worden.

Conclusie:

De tijd om tot een (gedeeltelijk) juiste associatie te komen lijkt voor de symbolen voor Internal Goal en Value Stream sneller te zijn in de nieuwe notatie dan voor de symbolen voor Value Proposition, Competence en Customer Goal.

Hoewel er uit de data blijkt dat er verschillen blijken bestaan tussen de gebruikte tijd voor PGA-1 en PGA-2 is significantie van deze verschillen niet aan te tonen. Voorafgaand aan het onderzoek was geen enkele van de respondenten bekend met de PGA-taal. Met andere woorden. Er kan niet aangetoond worden dat PGA-2 leidt tot een snellere (gedeeltelijk) juiste associatie. Deze conclusie wijkt niet af van de conclusie in de dataset van de individuele onderzoeker.

Comprehension Questions

Ranks				
	VersionCode	N	Mean Rank	Sum of Ranks
CQQ1Code	0	24	23.50	564.00
	1	20	21.30	426.00
	Total	44		
CQQ2Code	0	24	26.25	630.00
	1	20	18.00	360.00
	Total	44		
CQQ3Code	0	24	22.58	542.00
	1	20	22.40	448.00
	Total	44		
CQQ4Code	0	24	22.33	536.00
	1	20	22.70	454.00
	Total	44		
CQQ5Code	0	24	21.42	514.00
	1	20	23.80	476.00
	Total	44		
CQQ6Code	0	24	22.50	540.00
	1	20	22.50	450.00
	Total	44		

Test Statistics ^a						
	CQQ1Code	CQQ2Code	CQQ3Code	CQQ4Code	CQQ5Code	CQQ6Code
Mann-Whitney U	216.000	150.000	238.000	236.000	214.000	240.000
Wilcoxon W	426.000	360.000	448.000	536.000	514.000	450.000
Z	-1.568	-2.459	-.131	-.111	-.967	.000
Asymp. Sig. (2-tailed)	.117	.014	.896	.912	.333	1.000

a. Grouping Variable: VersionCode

- H0 Comprehension Questions: Score PGA-1 = Score PGA-2
- H1 Comprehension Questions: Score PGA-2 > Score PGA-1

Q1: Customer Goal

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.117 → 1-tailed sig = $(1 - (0.117 / 2)) = 0.9415 > 0,05$ H0 → kan niet verworpen worden

Q2 Value Proposition

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.014 → 1-tailed sig = $(1 - (0.014 / 2)) = 0.993 > 0,05$ H0 → kan niet verworpen worden

Q3: Value Stream

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.896 → 1-tailed sig = $(1 - (0,896 / 2)) = 0.552 > 0,05$ H0 → kan niet verworpen worden

Q4: Competence

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.912 → 1-tailed sig = $(1 - (0.912 / 2)) = 0.8335 > 0,05$ H0 → kan niet verworpen worden

Q5: Internal goal:

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.333 → 1-tailed sig = $(1 - (0.333 / 2)) = 0.8335 > 0,05$ H0 → kan niet verworpen worden

Q6: Importance *

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 = PGA-1

2-tailed Sig = 1.000 → 1-tailed sig = $(1 - (1.000 / 2)) = 0.500 > 0,05$ H0 → kan niet verworpen worden

*de case study is de enige opdracht waarin dit element werd gevraagd.

Conclusie:

In tegenstelling tot de dataset van de individuele onderzoeker blijkt dat de oude notatie hoger scoort voor alle symbolen. Hierbij moet wel worden aangetekend dat in het algemeen de verschillen marginaal zijn. Alleen het symbool voor Value proposition kent een goed waarneembaar verschil in score.

Het symbool voor importance scoort geen verschil in de oude en de nieuwe notatie. Deze is echter in geen enkele andere taak bevraagd. Om die reden weegt dit symbool minder zwaar in deze conclusies.

Hoewel de verschillen tussen de oude en de nieuwe notatie niet significant blijkt te zijn, is het beter scoren van de oude notatie en op zijn gunstigst gelijk scoren aan de nieuwe notatie wel een signaal dat voor het begrip van de symbolen in een taak de nieuwe notatie minder intuïtief lijkt te zijn. Voorafgaand aan de taak hebben de respondenten een summier uitleg van de PGA taal gehad.

Comprehension Questions Group time

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Group time: Comprehension questions - Q1	.211	44	.000	.724	44	.000
Group time: Comprehension questions - Q2	.197	44	.000	.757	44	.000
Group time: Comprehension questions - Q3	.334	44	.000	.380	44	.000
Group time: Comprehension questions - Q4	.199	44	.000	.881	44	.000
Group time: Comprehension questions - Q5	.265	44	.000	.610	44	.000
Group time: Comprehension questions - Q6	.213	44	.000	.658	44	.000

a. Lilliefors Significance Correction

Uit de normaliteitstest blijkt dat de significantie in alle gevallen kleiner is dan 0,05. Daarnaast is er sprake van twee groepen en van niet-gepaarde data.

Dit betekent dat er sprake is van een interval verdeling. Hiervoor wordt een Wilcoxon-Mann-Whitney test uitgevoerd.

- H0 Comprehension: Grouptime PGA-1 snelheid = Grouptime PGA-2 snelheid
- H1 Comprehension: Grouptime PGA-2 snelheid < Grouptime PGA-1 snelheid

Wilcoxon-Mann-Whitney Test

Ranks				
	VersionCode	N	Mean Rank	Sum of Ranks
Group time: Comprehension questions - Q1	0	24	22.13	531.00
	1	20	22.95	459.00
	Total	44		
Group time: Comprehension questions - Q2	0	24	22.67	544.00
	1	20	22.30	446.00
	Total	44		
Group time: Comprehension questions - Q3	0	24	20.69	496.50
	1	20	24.68	493.50
	Total	44		
Group time: Comprehension questions - Q4	0	24	20.00	480.00
	1	20	25.50	510.00
	Total	44		
Group time: Comprehension questions - Q5	0	24	21.13	507.00
	1	20	24.15	483.00
	Total	44		
Group time: Comprehension questions - Q6	0	24	27.17	652.00
	1	20	16.90	338.00
	Total	44		

Test Statistics ^a						
	Group time: Comprehension questions - Q1	Group time: Comprehension questions - Q2	Group time: Comprehension questions - Q3	Group time: Comprehension questions - Q4	Group time: Comprehension questions - Q5	Group time: Comprehension questions - Q6
Mann-Whitney U	231.000	236.000	196.500	180.000	207.000	128.000
Wilcoxon W	531.000	446.000	496.500	480.000	507.000	338.000
Z	-.212	-.094	-1.025	-1.414	-.778	-2.640
Asymp. Sig. (2-tailed)	.832	.925	.305	.157	.437	.008

a. Grouping Variable: VersionCode

Q1: Customer Goal

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 > PGA-1

2-tailed Sig = 0.832 → 1-tailed sig = $(1 - (0.832 / 2)) = 0.584 > 0,05 H_0 \rightarrow$ kan niet verworpen worden

Q2 Value Proposition

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.925 → 1-tailed sig = $0.925 / 2 = 0.4625 > 0,05$ H0 → kan niet verworpen worden

Q3: Value Stream

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 > PGA-1

2-tailed Sig = 0,305 → 1-tailed sig = $(1 - (0.305 / 2)) = 0.8475 > 0,05$ H0 → kan niet verworpen worden

Q4: Competence

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.157 → 1-tailed sig = $0.157 / 2 = 0.0785 > 0,05$ H0 → kan niet verworpen worden

Q5: Internal goal:

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.437 → 1-tailed sig = $0.221 / 2 = 0.2185 > 0,05$ H0 → kan niet verworpen worden

Q6: Importance *

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.008 → 1-tailed sig = $0.008 / 2 = 0.004 < 0,05$ H1 → kan aanvaard worden

Conclusie:

Hoewel er uit de data blijkt dat er verschillen bestaan in de besteedde tijd aan het oplossen van de casestudy tussen PGA-1 en PGA-2 is significantie van deze verschillen niet aan te tonen, met uitzondering van Q6. Hierbij zijn de verschillen tussen de groepen significant, waarmee gesteld kan worden dat voor Q6 met zekerheid geldt dat PGA-2 leidt tot een snellere oplossing van de taak. Het is echter niet zo dat dit betekent dat PGA-2 daarmee intuitiever is voor het symbool voor Importance. Uit de beantwoording van de vragen binnen de taak is gebleken dat dit symbool gelijk scoort in de oude en de nieuwe notatie.

Voorafgaand aan de taak hebben de respondenten een summiere uitleg van de PGA taal gehad.

Recall Questions

Wilcoxon-Mann-Whitney Test

Ranks				
	VersionCode	N	Mean Rank	Sum of Ranks
RQQ1Code	0	24	22.17	532.00
	1	20	22.90	458.00
	Total	44		
RQQ2Code	0	24	25.67	616.00
	1	20	18.70	374.00
	Total	44		
RQQ3Code	0	24	23.63	567.00
	1	20	21.15	423.00
	Total	44		
RQQ4Code	0	24	25.58	614.00
	1	20	18.80	376.00
	Total	44		
RQQ5Code	0	24	26.19	628.50
	1	20	18.08	361.50
	Total	44		

Test Statistics ^a					
	RQQ1Code	RQQ2Code	RQQ3Code	RQQ4Code	RQQ5Code
Mann-Whitney U	232.000	164.000	213.000	166.000	151.500
Wilcoxon W	532.000	374.000	423.000	376.000	361.500
Z	-.204	-1.956	-.729	-2.032	-2.240
Asymp. Sig. (2-tailed)	.838	.051	.466	.042	.025

a. Grouping Variable: VersionCode

- H0 Recall: Score PGA-1 = Score PGA-2
- H1 Recall: Score PGA-2 > Score PGA-1

Q1: Internal goal:

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.838 → 1-tailed sig = (1 – (0,838 / 2)) = 0.581 > 0,05 H0 → kan niet verworpen worden

Q2: Value Proposition

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.051 → 1-tailed sig = $(1 - (0,051 / 2)) = 0.9745 > 0,05 H_0 \rightarrow$ kan niet verworpen worden

Q3: Competence

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Score PGA-2 > PGA-1

2-tailed Sig = 0.466 → 1-tailed sig = $0.466 / 2 = 0.322 > 0,05 H_0 \rightarrow$ kan niet verworpen worden

Q4: Value Stream

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.042 → 1-tailed sig = $(1 - (0,042 / 2)) = 0.979 > 0,05 H_0 \rightarrow$ kan niet verworpen worden

Q5: Customer Goal

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.025 → 1-tailed sig = $(1 - (0.025/2)) = 0.9875 > 0,05 H_0 \rightarrow$ kan niet verworpen worden

Conclusie:

Hoewel er uit de data blijkt dat er verschillen bestaan in het herinneren van PGA-1 en PGA-2 is significantie van deze verschillen niet aan te tonen. In geval van de symbool voor Competence lijkt het zo te zijn dat deze beter herinnerd wordt met de nieuwe notatie. Dat dit symbool beter onthouden lijkt te worden blijkt ook uit de individuele dataset. Ondanks dat de verschillen in zowel de individuele als de gecombineerde dataset niet significant zijn is het kunnen bevestigen door te vergelijken met

een grotere dataset wel een indicatie dat in eventueel vervolgonderzoek PGA-2 beter zal scoren. Voorafgaand aan de taak hebben de respondenten gedurende een minuut de symbolen en hun betekenis in zich op kunnen nemen.

Recall Questions Group Time

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Group time: Recall questions - Q1	.187	44	.001	.707	44	.000
Group time: Recall questions - Q2	.176	44	.001	.850	44	.000
Group time: Recall questions - Q3	.181	44	.001	.819	44	.000
Group time: Recall questions - Q4	.225	44	.000	.642	44	.000
Group time: Recall questions - Q5	.177	44	.001	.853	44	.000

a. Lilliefors Significance Correction

Uit de normaliteitstest blijkt dat de significantie in alle gevallen kleiner is dan 0,05. Daarnaast is er sprake van twee groepen en van niet-gepaarde data. Dit betekent dat er sprake is van een interval verdeling. Hiervoor wordt een Wilcoxon-Mann-Whitney test uitgevoerd.

- H0 Recall: Grouptime PGA-1 snelheid = Grouptime PGA-2 snelheid
- H1 Recall: Grouptime PGA-2 snelheid < Grouptime PGA-1 snelheid

Wilcoxon-Mann-Whitney Test

Ranks				
	VersionCode	N	Mean Rank	Sum of Ranks
Group time: Recall questions - Q1	0	24	19.50	468.00
	1	20	26.10	522.00
	Total	44		
Group time: Recall questions - Q2	0	24	23.71	569.00
	1	20	21.05	421.00
	Total	44		
Group time: Recall questions - Q3	0	24	19.58	470.00
	1	20	26.00	520.00
	Total	44		
Group time: Recall questions - Q4	0	24	21.77	522.50
	1	20	23.38	467.50
	Total	44		
Group time: Recall questions - Q5	0	24	21.02	504.50
	1	20	24.28	485.50
	Total	44		

Test Statistics ^a					
	Group time: Recall questions - Q1	Group time: Recall questions - Q2	Group time: Recall questions - Q3	Group time: Recall questions - Q4	Group time: Recall questions - Q5
Mann-Whitney U	168.000	211.000	170.000	222.500	204.500
Wilcoxon W	468.000	421.000	470.000	522.500	504.500
Z	-1.697	-.684	-1.650	-.412	-.837
Asymp. Sig. (2-tailed)	.090	.494	.099	.680	.403

a. Grouping Variable: VersionCode

Q1: Internal goal:

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.090 → 1-tailed sig = 0.090 / 2 = **0.045** > 0,05 H1 → **kan aanvaard worden**

Q2: Value Proposition

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.494 → 1-tailed sig = 0,247 / 2 = 0.492 > 0,05 H0 → kan niet verworpen worden

Q3: Competence

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.099 → 1-tailed sig = 0.099 / 2 = **0.0495** > 0,05 H1 → **kan aanvaard worden**

Q4: Value Stream

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 > PGA-1

2-tailed Sig = 0.680 → 1-tailed sig = (1 – (0,680 / 2)) = 0.660 > 0,05 H0 → kan niet verworpen worden

Q5: Customer Goal

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 > PGA-1

2-tailed Sig = 0.403 → 1-tailed sig = $(1 - (0,403 / 2)) = 0.7985 > 0,05$ H0 → kan niet verworpen worden

Conclusie:

Uit de data blijkt dat, met uitzondering van de group time voor de symbolen voor Value Proposition en Competence, de tijd om PGA-2 te herinneren langer is dan de tijd om PGA-1 te herinneren. In geval van deze twee symbolen zijn de verschillen ook significant. Dit betekent dat hier gesteld kan worden dat toeval vrijwel zeker geen rol speelt in de tijd die nodig is om het symbool correct te herinneren.

De overige verschillen blijken echter niet significant. Langere tijd nodig om te herinneren duidt niet perse op minder intuïtief zijn van de notatie.

.

Representatievoorkeur

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
CustomerGoalCode	.117	44	.153	.942	44	.028
ValuePropositionCode	.157	44	.008	.896	44	.001
CompetenceCode	.179	44	.001	.930	44	.011
InternalGoalCode	.098	44	.200*	.964	44	.182
ValueStreamCode	.112	44	.200*	.921	44	.005

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Uit de normaliteitstest blijkt dat de significantie in groter is dan 0,05 voor het symbool voor Internal Goal. Voor dit symbool geldt dat er sprake is van een normale verdeling. Voor de overige symbolen is de sig. < 0.05. Hier geldt dat er sprake is van een niet-normale verdeling.

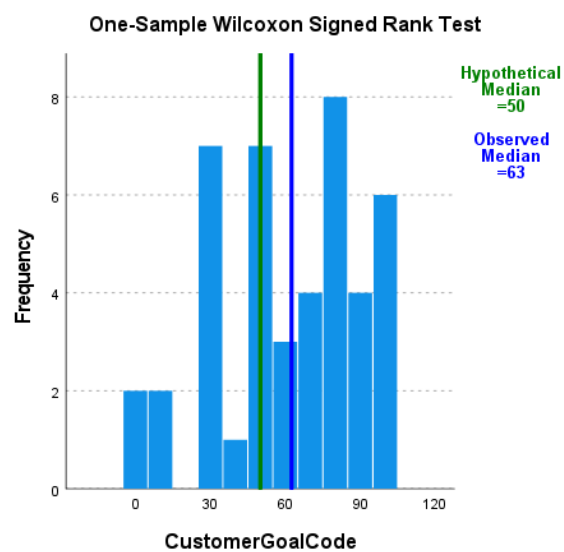
Gebruikmakend van onderstaande tabel volgt hieruit een t-Test voor gepaarde data voor het symbool voor Internal Goal, voor alle overige symbolen wordt de 1 sample Wilcoxon signed Rank Test uitgevoerd

Type of dependent variable	Type of independent variable						
	Ordinal/categorical				Normal/interval (ordinal)	More than 1	None
	Two groups		More groups				
	Paired	Unpaired	Paired	Unpaired			
2 categories	McNemar Test, Sign-Test	Fisher Test, Chi-squared-Test	Cochran's Q-Test	Fisher Test, Chi-squared-test	(Conditional) Logistic Regression	Logistic Regression	Chi-squared-Test
Nominal	Bowker Test	Fisher Test, Chi-squared-Test		Fisher Test, Chi-squared-test	Multinomial logistic regression	Multinomial logistic regression	Binomial Test
Ordinal	Wilcoxon Test, Sign-Test	Wilcoxon-Mann-Whitney Test	Friedman-Test	Kruskal-Wallis Test	Spearman-rank-test	Ordered logit	Median Test
Interval	Wilcoxon Test, Sign-Test	Wilcoxon-Mann-Whitney Test	Friedman-Test	Kruskal-Wallis Test	Spearman-rank test	Multivariate linear model	Median Test
Normal	t-Test (for paired)	t-Test (for unpaired)	Linear Model (ANOVA)	Linear Model (ANOVA)	Pearson-Correlation-test	Multivariate Linear Model	t-Test
Censored Interval	Log-Rank Test		Survival Analysis, Cox proportional hazards regression				
None	Clustering, factor analysis, PCA, canonical correlation						

CustomerGoalCode

One-Sample Wilcoxon Signed Rank Test Summary

Total N	44
Test Statistic	486.000
Standard Error	65.999
Standardized Test Statistic	2.038
Asymptotic Sig.(2-sided test)	.042



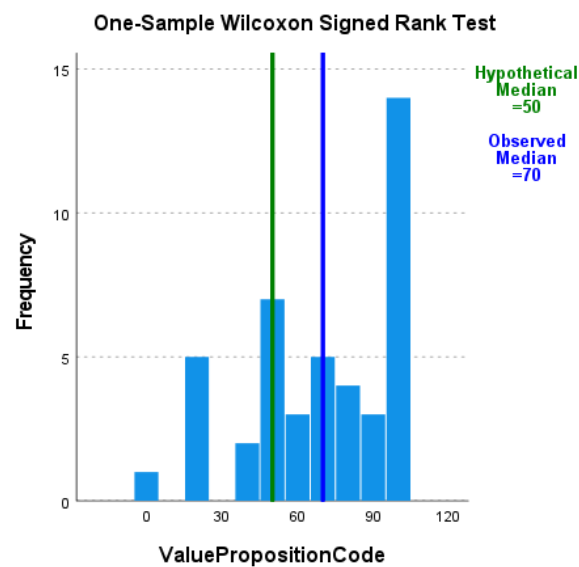
2 sided test sig: 0.042 → 1-sided sig = **0.042** / 2 = 0.021 < 0.05

Dit is **significant** en daarmee is er een duidelijke voorkeur voor de nieuwe notatie

ValuePropositionCode

One-Sample Wilcoxon Signed Rank Test Summary

Total N	44
Test Statistic	608.000
Standard Error	68.503
Standardized Test Statistic	3.467
Asymptotic Sig.(2-sided test)	.001



2 sided test sig: 0.001 \rightarrow 1-sided sig = $0.001 / 2 = 0.0005 < 0.05$

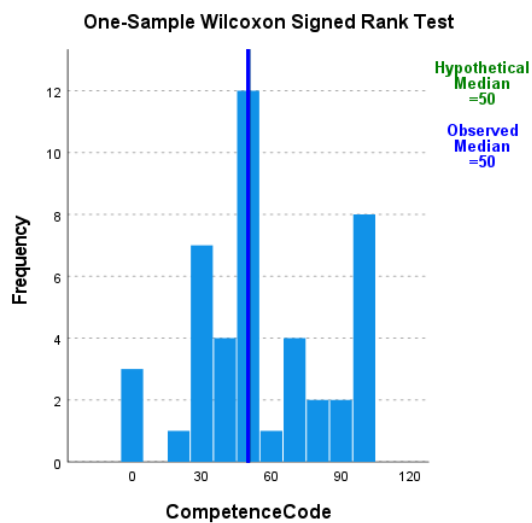
Dit **is significant** en daarmee is er een duidelijke voorkeur voor de nieuwe notatie

CompetenceCode

One-Sample Wilcoxon Signed Rank Test

Summary

Total N	44
Test Statistic	342.000
Standard Error	55.645
Standardized Test Statistic	1.105
Asymptotic Sig.(2-sided test)	.269



2 sided test sig: $0.269 \rightarrow 1\text{-sided sig} = 0.269 / 2 = 0.1345 > 0.05$

Dit is niet significant, waarmee de nulhypothese zoals onderstaand niet kan worden verworpen, hetgeen gezien de waargenomen mediaan van 50 in de lijn der verwachting ligt.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig. ^{a,b}	Decision
1	The median of CompetenceCode equals 50.	One-Sample Wilcoxon Signed Rank Test	.269	Retain the null hypothesis.

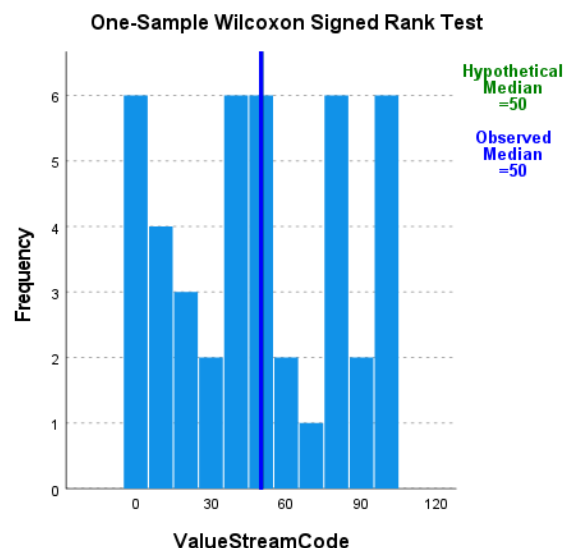
a. The significance level is .050.

b. Asymptotic significance is displayed.

ValueStreamCode

One-Sample Wilcoxon Signed Rank Test Summary

Total N	44
Test Statistic	335.000
Standard Error	68.649
Standardized Test Statistic	-.517
Asymptotic Sig.(2-sided test)	.605



2 sided test sig: $0.605 \rightarrow 1\text{-sided sig} = 0.605 / 2 = 0.3025 > 0.05$

Dit is niet significant, waarmee de nulhypothese zoals onderstaand niet kan worden verworpen, hetgeen gezien de waargenomen mediaan van 50 in de lijn der verwachting ligt.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig. ^{a,b}	Decision
1	The median of ValueStreamCode equals 50.	One-Sample Wilcoxon Signed Rank Test	.605	Retain the null hypothesis.

a. The significance level is .050.

b. Asymptotic significance is displayed.

T-Test

InternalGoalCode

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
InternalGoalCode	44	53.52	27.840	4.197

One-Sample Test

				Test Value = 50		
				95% Confidence Interval of the Difference		
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
InternalGoalCode	.839	43	.406	3.523	-4.94	11.99

De mediaan ligt net boven de 50.

2 tailed test sig: 0.406 → 1-sided sig = 0.406 / 2 = 0.203 < 0.05 Dit is niet significant.

Conclusie:

Over het algemeen lijkt er een voorkeur te bestaan voor de nieuwe notatie. Dit zou kunnen betekenen dat de nieuwe notatie als intuïtiever wordt ervaren, met uitzondering van het symbool voor competence. Hier lijkt geen voorkeur te bestaan.

Bijlage 6: Testen geaggregeerde scores

Geaggregeerde scores over alle symbolen heen individuele dataset

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Geaggregeerd Internal Goal	.164	15	.200*	.935	15	.325
Geaggregeerd Value proposition	.189	15	.158	.963	15	.737
Geaggregeerd Competence	.191	15	.145	.941	15	.393
Geaggregeerd Value Stream	.225	15	.040	.853	15	.019
Geaggregeerd Customer Goal	.107	15	.200*	.980	15	.969

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Met uitzondering van de resultaten voor het symbool Value Stream is de sig > 0.05. Dit houdt een niet-normale verdeling voor dit symbool en daarmee een Wilcoxon-Mann-Whitneytest.

Voor alle overige symbolen is er sprake van een interval verdeling en moet een t-Test worden uitgevoerd.

T-Test

Group Statistics					
	VersionCode	N	Mean	Std. Deviation	Std. Error Mean
Geaggregeerd Internal Goal	0	10	234.8889%	64.03381%	20.24927%
	1	5	238.0000%	58.39758%	26.11619%
Geaggregeerd Value proposition	0	10	201.0000%	51.59912%	16.31707%
	1	5	159.6667%	100.70860%	45.03826%
Geaggregeerd Competence	0	10	205.6667%	47.53946%	15.03330%
	1	5	211.3333%	78.15050%	34.94996%
Geaggregeerd Customer Goal	0	10	256.1667%	38.80189%	12.27023%
	1	5	269.3333%	41.19061%	18.42100%

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Geaggregeerd Internal Goal	Equal variances assumed	.061	.808	-.091	13	.929	-3.111111%	34.15262%	-76.89337%	70.67115%
	Equal variances not assumed			-.094	8.836	.927	-3.111111%	33.04676%	-78.08044%	71.85822%
Geaggregeerd Value proposition	Equal variances assumed	1.418	.255	1.071	13	.304	41.333333%	38.58989%	-42.03505%	124.70172%
	Equal variances not assumed			.863	5.080	.427	41.333333%	47.90294%	-81.22361%	163.89028%
Geaggregeerd Competence	Equal variances assumed	.520	.484	-.176	13	.863	-5.66667%	32.14271%	-75.10677%	63.77344%
	Equal variances not assumed			-.149	5.533	.887	-5.66667%	38.04602%	-100.70067%	89.36734%
Geaggregeerd Customer Goal	Equal variances assumed	.065	.802	-.608	13	.554	-13.16667%	21.66366%	-59.96816%	33.63482%
	Equal variances not assumed			-.595	7.666	.569	-13.16667%	22.13350%	-64.59603%	38.26270%

In alle gevallen is er sprake van een significantie > 0,05 in de Levene's test. Dit betekent dat er kan worden uitgegaan van gelijke varianties.

Internal goal:

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.929 → 1-tailed sig = $(1 - (0.929 / 2)) = 0,5355 > 0,05$ → Verschillen zijn niet significant

Value Proposition

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.304 → 1-tailed sig = $(1 - (0.304 / 2)) = 0,848 > 0,05$ → Verschillen zijn niet significant

Competence

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.863 → 1-tailed sig = $(1 - (0.863 / 2)) = 0.5685 > 0,05$ → Verschillen zijn niet significant

Customer Goal

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 > PGA-1

2-tailed Sig = 0.554 → 1-tailed sig = $0.554/2 = 0.277 > 0,05$ → Verschillen zijn niet significant

Wilcoxon-Mann-Whitneytest

Ranks				
	VersionCode	N	Mean Rank	Sum of Ranks
Geaggregeerd Value Stream	0	10	8.80	88.00
	1	5	6.40	32.00
	Total	15		

Test Statistics^a

Geaggregeerd Value Stream	
Mann-Whitney U	17.000
Wilcoxon W	32.000
Z	-.981
Asymp. Sig. (2-tailed)	.327
Exact Sig. [2*(1-tailed Sig.)]	.371 ^b

a. Grouping Variable: VersionCode

b. Not corrected for ties.

Value Stream

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.327 → 1-tailed sig = $(1 - (0.327 / 2)) = 0.8365 > 0,05$ → Verschillen zijn niet significant

Geaggregeerde scores over alle symbolen heen gecombineerde dataset

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Geaggregeerd Internal Goal	.171	44	.002	.882	44	.000
Geaggregeerd Value proposition	.118	44	.141	.978	44	.559
Geaggregeerd Competence	.061	44	.200*	.984	44	.776
Geaggregeerd Value Stream	.170	44	.003	.936	44	.016
Geaggregeerd Customer Goal	.160	44	.007	.929	44	.010

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

De resultaten voor de symbolen Value Proposition en Competence hebben een sig > 0.05. Dit houdt een normale verdeling voor deze symbool en moet een t-Test worden uitgevoerd.

Voor alle overige symbolen is sig < 0.05 en is er sprake van een niet-normale verdeling waarmee een Wilcoxon-Mann-Whitneytest moet worden uitgevoerd voor deze symbolen

Wilcoxon-Mann-Whitneytest

Ranks				
	VersionCode	N	Mean Rank	Sum of Ranks
Geaggregeerd Internal Goal	0	24	21.79	523.00
	1	20	23.35	467.00
	Total	44		
Geaggregeerd Value Stream	0	24	24.73	593.50
	1	20	19.83	396.50
	Total	44		
Geaggregeerd Customer Goal	0	24	23.08	554.00
	1	20	21.80	436.00
	Total	44		

Test Statistics ^a			
	Geaggregeerd Internal Goal	Geaggregeerd Value Stream	Geaggregeerd Customer Goal
Mann-Whitney U	223.000	186.500	226.000
Wilcoxon W	523.000	396.500	436.000
Z	-.401	-1.262	-.330
Asymp. Sig. (2-tailed)	.689	.207	.741

a. Grouping Variable: VersionCode

Internal goal:

Groep A (0) Nieuwe notatie (PGA-2)

Groep B (1) Oude notatie (PGA-1)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig =0.689 → 1-tailed sig = (1 - (0.689 / 2)) = 0,6555 > 0,05 H0 → Verschillen zijn niet significant

Value Stream

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 < PGA-1

2-tailed Sig = 0.207 → 1-tailed sig = $(1 - (0.207 / 2)) = 0,8965 > 0,05$ H0 → Verschillen zijn niet significant

Customer Goal

Groep A (0) Oude notatie (PGA-1)

Groep B (1) Nieuwe notatie (PGA-2)

Ranking: Score PGA-2 > PGA-1

2-tailed Sig = 0.741 → 1-tailed sig = $0.741 / 2 = 0.6295 > 0,05$ H0 → Verschillen zijn niet significant

Independant Samples t-Test

Group Statistics					
	VersionCode	N	Mean	Std. Deviation	Std. Error Mean
Geaggregeerd Value proposition	0	24	201.7361%	61.92461%	12.64031%
	1	20	180.1667%	88.92837%	19.88499%
Geaggregeerd Competence	0	24	214.5602%	71.37067%	14.56848%
	1	20	195.8889%	71.87080%	16.07080%

Independent Samples Test										
Levene's Test for Equality of Variances				t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Geaggregeerd Value proposition	Equal variances assumed	4.539	.039	.945	42	.350	21.56944%	22.81305%	-24.46915%	67.60804%
	Equal variances not assumed			.915	33.005	.367	21.56944%	23.56247%	-26.36847%	69.50736%
Geaggregeerd Competence	Equal variances assumed	.046	.832	.861	42	.394	18.67130%	21.67718%	-25.07502%	62.41761%
	Equal variances not assumed			.861	40.477	.394	18.67130%	21.69127%	-25.15229%	62.49489%

Value Proposition

Groep A (0) Oude notatie (PGA-1)
Groep B (1) Nieuwe notatie (PGA-2)
Ranking: Score PGA-2 < PGA-1
Levene's sig: 0.039 -> geen gelijke varianties
2-tailed Sig = 0.367 → 1-tailed sig = (1 – (0,367 / 2)) = 0.8165 > 0,05 H0 → Verschillen zijn niet significant

Competence

Groep A (0) Nieuwe notatie (PGA-2)
Groep B (1) Oude notatie (PGA-1)
Ranking: Score PGA-2 < PGA-1
Levene's sig: 0.832 -> gelijke varianties
2-tailed Sig = 0.394 → 1-tailed sig = (1 – (0.394 / 2)) = 0.803 > 0,05 H0 → Verschillen zijn niet significa

Conclusies:

Wanneer we de individuele dataset met de gecombineerde dataset vergelijken zien we exact dezelfde resultaten. Alleen in het geval van het symbool voor Customer Goal lijkt een voorkeur te bestaan voor PGA-2. In de overige gevallen lijkt de voorkeur te liggen bij PGA-1. De verschillen zijn echter eveneens in beide gevallen niet significant.

